A NOVEL NUCLEIC ACID MOLECULE

FIELD OF THE INVENTION

The present invention is directed generally to an isolated nucleic acid molecule encompassing a neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof and its use *inter alia* in developing a range of eukaryotic artificial chromosomes including mammalian (e.g. human) and non-mammalian artificial chromosomes. Such artificial chromosomes are useful in a range of genetic therapies.

BACKGROUND OF THE INVENTION

Bibliographic details of the publications referred to by author in this specification are collected at the end of the description.

The rapidly increasing sophistication of recombinant DNA technology is greatly facilitating research and development in the medical and allied health fields. A particularly important area is in mammalian including human genetics and the molecular mechanisms behind some genetic abnormalities. Progress in research in this area has been hampered by the lack of a cloned nucleic acid molecule encompassing a human centromere. The identification and cloning of a human centromere will promote the development of techniques for introducing genes into eukaryotic cells and in particular mammalian including human cells and will be an important asset to gene therapy and the development of a range of genetic diagnostic tests.

The centromere is an essential structure for sister chromatid cohesion and proper chromosomal segregation during mitotic and meiotic cell divisions. The centromere of the budding yeast *Saccharomyces cerevisiae* has been extensively studied and shown to be contained within a relatively short DNA segment of 125 bp that is organized into an 8-bp (CDEI) and 26-bp (CDEIII) domain, separated by a 78- to 87-bp, highly AT-rich, middle (CDEII) domain (Clarke and Carbon, 1985). The centromere of the fission yeast *Schizosaccharomyces pombe* is considerably larger, ranging from 40 to 100 kb, and consists of a central core DNA element of 4 to 7 kb flanked on both sides by inverted repeat units (Steiner *et al.*, 1993). Recently, the

functional DNA components of a higher eukaryotic centromere have been characterized in a minichromosome from *Drosophila melanogaster* and shown to consist of a 220-kb essential core DNA flanked by 200 kb of highly repeated sequences on one side (Murphy and Karpen, 1995).

The mammalian centromere, like the centromeres of all higher eukaryotes studied to date, contains a great abundance of highly repetitive, heterochromatic DNA. For example, a typical human centromere contains 2 to 4 Mb of the 171-bp α-satellite repeat (Wevrick and Willard 1989, 1991; Trowell *et al.*, 1993), plus a smaller and more variable quantity of a 5-bp satellite III DNA (Grady *et al.*, 1992; Trowell *et al.*, 1993). The role of these satellite sequences is presently unclear. Transfection of a cloned 17-kb uninterrupted α-satellite array into cultured simian cells (Haaf *et al.*, 1992) or a 120-kb α-satellite-containing YAC into human and hamster cells (Larin *et al.*, 1994) appear to confer centromere function at the sites of integration. Other workers have analyzed rearranged Y chromosomes (Tyler-Smith *et al.*, 1993), or dissected the centromere of the human Y chromosome with cloned telomeric DNA (Brown *et al.*, 1994) and suggested that 150 to 200 kb of α-satellite DNA plus ~300 kb of adjacent sequences are associated with human centromere function. In addition, a human X-derived minichromosome that retained 2.5 Mb of α-satellite array has been produced by telomere-associated chromosome fragmentation (Farr *et al.*, 1995). In all these studies, it is not known whether non-α-satellite DNA sequences are embedded within the centromeric site and operate independently of, or in concert with, the α-satellite DNA.

In mammals, four constitutive centromere-binding proteins, CENP-A, CENP-B, CENP-C, and CENP-D, have been characterized to varying extents and implicated to have possible direct roles in centromere function. CENP-A, a protein localized to the outer kinetochore domain, is a centromere-specific core histone that shows sequence homology to the histone H3 protein and may serve to differentiate the centromere from the rest of the chromosome at the most fundamental level of chromatin structure - the nucleosome (Sullivan *et al.*, 1994). CENP-B, a protein which associates with the centromeric heterochromatin through its binding to the CENP-B box motif found in primate α-satellite and mouse minor satellite DNA, probably has a role in packaging centromeric heterochromatic DNA - a role which, however, may not be indispensable since the protein is undetectable on the Y chromosome (Pluta *et al.*, 1990) and is found on the

inactive centromeres of dicentric chromosomes (Earnshaw *et al.*, 1989). CENP-C has been shown to be located at the inner kinetochore plate and is postulated to have an essential although yet undetermined centromere function, as seen, for example, from inhibition of mitotic progression following microinjection of anti-CENP-C antibodies into cells (Bernat *et al.*, 1990; Tomkiel *et al.*, 1994) and from its association with the active but not the inactive centromeres of dicentric chromosomes (Earnshaw *et al.*, 1989; Page *et al.*, 1995; Sullivan and Schwartz, 1995). Finally, CENP-D (or RCC1) is a guanine exchange factor that appears to have a general cellular role that is neither specific nor clear for the centromere (Kingwell and Rattner 1987; Bischoff *et al.*, 1990; Dasso, 1993). More recently, a new role for the mammalian centromere as a "marshalling station" for a host of "passenger proteins" (such as INCENPs, MCAK, CENP-E, CENP-F, 3F3/2 antigens, and cytoplasmic dynein), has been recognized (reviewed by Earnshaw and Mackay, 1994, and Pluta *et al.*, 1995). These passenger proteins, whose appearance at the centromere is transient and tightly regulated by the cell cycle, provide vital functions that include motor movement of chromosomes, modulation of spindle dynamics, nuclear organization, intercellular bridge structure and function, sister chromatid cohesion and release, and cytokinesis. At present, except for CENP-B, none of the constitutive or passenger proteins have been demonstrated to bind mammalian centromere DNA directly.

In work leading up to the present invention, the inventors identified in a patient (hereinafter referred to as "BE") an unusual human marker chromosome, mardel 10, which is 100% stable in mitotic division both in patient BE and in established fibroblast and transformed lymphoblast cultures. In accordance with the present invention, a region of the mardel (10) chromosome has been cloned together with the corresponding region from a normal human subject. The nucleic acid molecules cloned contain no substantial α-satellite repeats yet are mitotically stable. The nucleic acid molecules encompass therefore, a new form of centromere referred to herein as a "neocentromere". The identification and cloning of a eukaryotic neocentromere without substantial α-satellite DNA repeat sequences now provides the means of generating a range of eukaryotic artificial chromosomes such as mammalian including human artificial chromosomes with uses in genetic therapy, transgenic plant and animal production and recombinant protein production. A range of diagnostic reagents is now also obtainable using the cloned neocentromere.

SUMMARY OF THE INVENTION

Sequence Identity Numbers (SEQ ID NOs.) for the nucleotide sequences referred to in the specification are defined following the bibliography.

Throughout this specification and the claims which follow, unless the context requires otherwise, the word "comprise", or variations such as "comprises" or "comprising", will be understood to imply the inclusion of a stated integer or group of integers but not the exclusion of any other integer or group of integers.

A fibroblast cell line 920158 carrying the mardel marker chromosome was deposited at the European Collection of Cell Cultures (ECACC), Centre for Applied Microbiology Research, Salisbury, Wiltshire, SP4 0JG, UK on 1 May, 1997 under Accession No. 97051716. Bacterial artificial chromosomes (BACs) carrying portions of the mardel (10) chromosome have also been deposited at ECACC as follows:

BAC/E8-1: deposited on 5 May 1998 under Accession Number 980505016;
BAC/F2-14: deposited on 5 May 1998 under Accession Number 980505017.

A number of human fibrosarcoma cell lines carrying various neocentromeric constructs were deposited at ECACC as described hereafter by Accession Number with the date of deposit in parenthesises.

HT-38	98050704 (7 May 1998)
HT-47	98050705 (7 May 1998)
HT-54	98050706 (7 May 1998)
HT-190	98050707 (7 May 1998)
HT-191	98050708 (7 May 1998).

One aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides derived from a eukaryotic chromosome and encompassing a

neocentromere or a functional derivative synthetic or hybrid form thereof which nucleic acid molecule or its derivatives, synthetic forms or hybrid forms when introduced into a compatible cell is capable of replicating, acting as an extra-chromosomal element and segregating with cell division.

Another aspect of the present invention contemplates a nucleic acid molecule or its chemical equivalent having a tertiary structure which defines a human neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof or its mammalian or non-mammalian homologue.

Yet a further aspect of the present invention is directed to an isolated nucleic acid molecule comprising a sequence of nucleotides encompassing a neocentromere derived from a eukaryotic chromosome, which nucleic acid molecule when introduced into a compatible cell is a replicating, extra-chromosomal element which segregates with cell division.

Still another aspect of the present invention is directed to an isolated nucleic acid molecule having a sequence of nucleotides or their chemical equivalents which directs a conformation defining a human neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof or a mammalian or non-mammalian homologue thereof wherein the neocentromere associates with centromere binding proteins (CENP) -A and CENP-C or antibodies thereto and does not contain substantial α-satellite DNA repeat sequences.

A further aspect of the present invention is directed to an isolated nucleic acid molecule comprising a nucleotide sequence encompassing a neocentromere or a functional derivative, synthetic or hybrid form thereof which when said nucleic acid molecule is in linear form and co-introduced into a cell together with a telomeric sequence, is capable of replicating, remaining as an extra-chromosomal element and segregates with cell division.

Another aspect of the present invention provides an isolated nucleic acid molecule or a derivative, synthetic or hybrid form thereof comprising a sequence of nucleotides:

(i) which directs conformation defining a human neocentromere or a functional derivative

thereof or a latent, synthetic or hybrid form thereof or its mammalian or non-mammalian homologue wherein said neocentromere is capable of associating with CENP-A and CENP-C;

(ii) which contains no substantial α-satellite DNA sequence repeat; and

(iii) which is capable, when introduced into compatible cells, of replication, remaining extra-chromosomal and segregating with cell division.

Even yet another aspect of the present invention is directed to a genetic construct comprising an origin of replication for a eukaryotic cell and a nucleic acid molecule encompassing a human neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof or its mammalian or non-mammalian homologue flanked by telomeric nucleotide sequences functional in the cell in which the genetic construct is to replicate and wherein said genetic construct when introduced into a cell is a replicating, extra-chromosomal element which segregates with cell division.

Another aspect of the present invention is directed to a genetic construct in the form of a eukaryotic artificial chromosome such as a mammalian artificial chromosome (MAC), a human artificial chromosome (HAC) or comprising an origin of replication and a sequence of nucleotides which:

(i) directs a conformation defining a human neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof wherein said neocentromere is capable of associating with CENP-A and CENP-C or antibodies thereto; and

(ii) contains no substantial α-satellite DNA repeat sequences;

said sequence of nucleotides flanked by eukaryotic (e.g. mammalian) telomeric sequences which are in turn flanked by yeast telomeric sequences wherein a unique enzyme site is located between the human and yeast telomeric nucleotide sequences such that upon contact with said enzyme, the yeast telomeric sequences are removed and the eukaryotic (e.g. mammalian) telomeric sequences are exposed.

Still another aspect of the present invention provides a genetic construct comprising an origin of replication and a first nucleic acid molecule defining a human neocentromere or a functional

derivative thereof or latent, synthetic or hybrid form thereof, a second nucleic acid molecule encoding a peptide, polypeptide or protein, wherein said first and second nucleic acid molecules are flanked by a first set of eukaryotic (e.g. mammalian, such as human) telomeric sequences which are in turn flanked by a second set of eukaryotic (e.g. yeast) telomeric sequences wherein there are unique enzyme sites between the first and second telomeric sequences such that upon contact with a required enzyme, the second telomeric sequences are cleaved off to expose the first telomeric sequences.

BRIEF DESCRIPTION OF THE FIGURES

Figure 1 is a schematic representation showing identification of a YAC contig spanning the marker centromere region. (A) Comparison of GTL banding patterns of mardel 10 and normal chromosome 10. The pair of open arrows indicate the two breakpoints on a normal chromosome 10 in generating the marker chromosome (Voullaire *et al.*, 1993). The long and short arms of the marker chromosome are designated q' and p', respectively, to distinguish them from the q and p arms of the normal chromosome 10. Asterisk denotes the position of a cosmid 10pC38 that was used to "tag" the q'-arm of stretched marker chromosomes in the ANTI-CEN/FISH experiments. (B) A 4-megabase YAC contig (#082) from 10q25.2 region that spans the marker centromere. The tilling path of YACs #0 to #23 and their corresponding CEPH library addresses are shown. (C) FISH mapping of selected YAC clones from contig #082 using normal fluorescence microscopy and standard metaphase chromosomes prepared from transformed lymphoblast cells of patient BE. The distribution of FISH signals (vertical axis) is shown as a percentage of the signals on one arm of the marker chromosome that is in excess of those found on the opposite arm of the chromosome. The total number of fluorescence signals scored for each of the YAC clones is indicated in brackets.

Figure 2 is a photographic representation showing ANTI-CEN/FISH analysis of the marker centromere. (A) Detection of α-satellite DNA using a mixture of α-satellite DNA probes (red signals) under low stringency conditions. Centromeres were counter-labelled with CREST#6 autoimmune antibody (pale blue dots; or white when superimposed on a red background). Chromosomes were prepared from transformed lymphoblast cells of patient BE. The right-hand

panel represents green pseudo-coloring of DAPI images of chromosomes to provide a better definition of chromosome outline. Only the signal for the antibody, but not that for α-satellite, was seen on the marker centromere (arrowed). (B) Simultaneous labelling of stretched human metaphase chromosomes with CREST#6 (red) and anti-CENP-C antibody, Am-C1 (pale blue), with the white color indicating full coincidence of the two antibody signals. (C) Detection of CENP-C on the marker chromosome. Simultaneous labelling of the marker chromosome (arrowhead) with (a) Am-C1 (pale blue) and (b) CREST#6 (red). (c) Combined images of a and b, showing complete coincidence of Am-C1 and CREST#6 signals. (d) FISH analysis of the same cell as a-c using the 10pC38 cosmid probe (pale blue dots and green arrows) to identify the marker chromosome. Some loss of ANTI-CEN signal, especially for the Am-C1 antibody was seen following FISH. (e) Green pseudo-coloring of DAPI images. A colour photograph corresponding to this figure is available upon request.

Figure 3 is a photographic representation showing ANTI-CEN/FISH analysis of cosmid clones on stretched (A, a-f) and superstretched (B) metaphase chromosomes. (a-c) Examples of cosmid signals (white arrows) localized to the q'-region of the marker centromere. (d-f) Examples of cosmid signals (white arrows) localized to the p'-region of the marker centromere. Green arrows indicate positions of the 10pC38 cosmid DNA tag used to mark the q'-end of the marker chromosome. (B) Mapping of Y6C21 onto a superstretched metaphase chromosome. Not included is the 10pC38 q'-tag signal located further to the left of the chromosomal segment shown. ANTI-CEN signals are in red, FISH signals are in pale blue, and overlapping ANTI-CEN and FISH signals are in white. Each of the pictures is accompanied by DAPI images of chromosomes pseudo-coloured in green. A colour photograph corresponding to this figure is available upon request.

Figure 4 Localization of the anti-centromere antibody-binding domain. ***a***, Relative positions of different cosmid and PAC clones within the YAC #082 contig, using YAC-3 as a reference. Cosmids are designated as YnCm, where 'n' denotes the YAC of origin and 'm' denotes the cosmid number. PACs 1-5 are five different PAC clones isolated from a human PAC library (Genome Systems Inc). "HC-contig" represents a group of overlapping cosmids that map tightly around the marker centromere in ANTI-CEN/FISH experiments. A genomic map corresponding

to the depicted YAC region was derived from the DNA of patient BE and shown above the YAC map. S, *Sal*I; K, *Ksp*I; N, *Not*I; Sf, *Sfi*I. ***b***, Cumulative scoring of FISH signals in ANTI-CEN/FISH experiments for cosmids Y3C64, Y6C8, Y3C94, Y7C14, Y4C45, Y6C10, Y6C21, Y3C3, PAC5, Y13C1, Y13C8, and Y17C6. The distribution of FISH signals (vertical axis) is those found on the opposite arm of the chromosome. The total number of fluorescence signals scored for each of the cosmid clones is indicated in brackets. ***c***, Restriction mapping of the 80-kb region covered by the eight overlapping cosmids of the HC-contig. These eight cosmids were derived from four different YACs (YAC-3, YAC-4, YAC-6, and YAC-7) and provided independent confirmation of the map. Furthermore, the map agreed fully with the restriction map of a 120 kb-insert PAC clone (PAC4) that spanned the entire HC-contig region. E, *Eco*RI; R, *Eco*RV; N, *Not*I.

Figure 5 is a representation showing restriction analysis of genomic DNA of patient BE and those of his normal parents using Y6C10 as probe. DNA was resolved on a PFGE (A) or standard agarose gel (B and C). Samples 1, 2 and 3 were fibroblast cultures of mother of BE, father of BE, and patient BE, respectively. Sample 4 was a somatic hybrid cell line BE2C1-18-5F containing the marker chromosome. Fragment sizes are in kilobases.

Figure 6 is a representation of the full nucleotide sequence of the HC-contig DNA derived from normal human chromosome 10q 25.2 region.

Figure 7 is a diagrammatic representation of the method used to retrofit YAC3 and YAC5.

Figures 8A to J are diagrammatic representations of the different vectors used for cloning DNA as YACs by the conventional restriction/ligation methods.

Figures 9 is a diagrammatic representation of circular TAR summarising the recombination process.

Figure 10 is a diagrammatic representation showing modification of TAR vector.

Figure 11 is a diagrammatic representation of the cloning of 10q25 human neocentromere DNA from mardel (10) chromosome. This DNA is designated NC-contig DNA to distinguish it from the HC-contig derived from the corresponding region of the normal chromosome 10. (A) Structural map of the NC-contig region and flanking DNA. Arrows indicate the relative positions and directions of primers used in PCR analyses (Table 3). The restriction sites *EcoR*I, *EcoR*V, *Srf*I, and *Sfi*I and *Sfi*I are indicated by E, R, Sr and Sf, respectively. The position of the TAR "hook" CE-F2 is represented by the solid box. The hatched bar represents HC- or NC-contig. p′ and q′ refer to the short and long arms of mardel (10), respectively. (B) Circular TAR strategy using the vectors pVC39-Alu/C3-F2(+) and pVC39-Alu/C3-F2(-) for the direct cloning of the neocentromere DNA from mardel (10). The position of the Alu consensus sequence hook is represented by the white box. Crosses denote the sites of recombination between the TAR vector and the genomic DNA at the Alu and C3-F2 hooks during cloning. (C) Structural maps of the resulting circular YACs 5f-52-E8 and 5f-38-F2 containing the neocentromere DNA of the mardel (10) chromosome. The DNA flanking the NC-contig is represented by stippled bars. (D) Structural maps of BAC/E8-1 and BAC/F2-14. Nt represents *Not*I and URA-BAC-neo represents the retrofitting vector BRV1 (Larionov *et al.*, 1997).

Figure 12 is a diagrammatic representation showing specific TAR of HC-region from mardel 10.

The method was as follows: (1) Co-transformation into YPH857; (2) Select HIS^+ colonies; (3) screen for HC-region by PCR; (4) Prepare high-MW DNA; (5) Digest with I-Sce1 to expose hTELS; (6) Transfect HT 1080 cells; (7) Select for $G418^R$; and (8) analyse by PFGE and FISH.

Figure 13 is a diagrammatic representation showing cloning in yeast as YAC/HAC.

Figure 14 is a diagrammatic representation outlining TACT procedure.

Figure 15 is a diagrammatic representation of TACT constructs.

Figure 16A is a representation of the full nucleotide sequence of the NC-contig DNA derived

from mardel (10) and corresponds to the HC-contig DNA region of the normal chromosome 10.

Figure 16B is a representation of the partial nucleotide sequence of the BAC/F2-14 clone that is derived from a region immediately p' of the NC-contig DNA (see Fig. 11D).

5

SUMMARY OF SEQ ID NOs.

SEQ ID NO.	DESCRIPTION
1	DNA primer
2	DNA primer
3	Nucleotide sequence of HC-contig
4	Nucleotide sequence of NC-contig
5	BAC-F2 contig 1
6	BAC-F2 contig 2
7	BAC-F2 contig 3
8	BAC-F2 contig 4
9	BAC-F2 contig 5
10	BAC-F2 contig 6
11	BAC-F2 contig 7
12	BAC-F2 contig 8
13	BAC-F2 contig 9
14	BAC-F2 contig 15
15	BAC-F2 contig 33
16	BAC-F2 contig 39
17	BAC-F2 contig 41
18	BAC-F2 contig 42
19	BAC-F2 contig 44
20	BAC-F2 contig 47
21	BAC-F2 contig 47 fragment 1
22	BAC-F2 contig 47 fragment 2
23	BAC-F2 contig 47 fragment 3
24	BAC-F2 contig 47 fragment 4
25	BAC-F2 contig 47 fragment 5
26	BAC-F2 contig 47 fragment 6

27	BAC-F2 contig 47 fragment 7
28	BAC-F2 contig 47 fragment 8
29	BAC-F2 contig 47 fragment 9

ABBREVIATIONS USED IN THE SUBJECT SPECIFICATION

mardel (10):	Marker chromosome from patient BE; comprises a rearrangement of chromosome 10.
HAC:	Human artificial chromosome
YAC:	Yeast artificial chromosome
MAC:	Bacterial artificial chromosome
PLAC:	Plant artificial chromosome
neocentromere:	A centromere containing no substantial α-satellite DNA
CENP:	Centromere binding protein
HC-contig:	Region of normal chromosome 10 comprising neocentromere
E8:	q' end/region of mardel (10) neocentromere
F2:	p' end/region of mardel (10) neocentromere
BE:	Patient from which mardel (10) identified
TAR:	Transformation-associated recombinant
PCR:	Polymerase chain reaction
Marker neocentromere:	neocentromere on mardel (10).
NC-contig	region of mardel (10) chromosome comprising neocentromere

DETAILED DESCRIPTION OF PREFERRED EMBODIMENTS

The present invention is predicated in part on the identification and isolation of nucleic acid molecules exhibiting neocentromeric properties. In accordance with the present invention, a neocentromere is considered a centromere which does not contain substantial α-satellite DNA repeat sequences and, when activated, is capable of functioning as a centromere. The term

"substantial" in this context means that the nucleic acid molecule does not contain detectable α-satellite by FISH analysis under medium stringency conditions. The neocentromere may contain a small number of highly diversed α-satellite DNA. In primates, α-satellite DNA is consider 171bph in length. An nucleic acid molecule containing an activated neocentromere or a neocentromere otherwise functioning as a centromere facilitates in accordance with the present invention, the nucleic acid molecule replicating, remaining extra-chromosomal and segregating with cell division. Reference herein to "neocentromere" is taken to mean a centromere substantially devoid of α-satellite DNA repeat sequences.

Accordingly, one aspect of the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides which defines an eukaryotic neocentromere.

More particularly the present invention provides an isolated nucleic acid molecule comprising a sequence of nucleotides derived from a eukaryotic chromosome and encompassing a neocentromere which nucleic acid molecule when introduced into a compatible cell is capable of replicating, acting as an extra-chromosomal element and segregating with cell division.

The present invention is exemplified herein by the identification and cloning of a human neocentromere. This is done, however, with the understanding that the present invention extends to all eukaryotic neocentromeres such as from mammalian, plant, aviary, insect, fungal, yeast and reptilian chromosomes. The most preferred neocentromere, however, is from human chromosomes and their mammalian homologues.

The present invention is predicated in part on the identification of an unusual chromosomal marker in a patient designated "BE". The chromosomal marker is referred to as "mardel (10)" and results from a rearrangement of human chromosome 10. The mardel (10) marker is mitotically stable and, in accordance with the present invention, contains a functional neocentromere at a location regarded as non-centromeric. The neocentromere at mardel (10) is located between q24 and q26 on chromosome 10 and more particularly around q25. Even more particularly, the neocentromere maps to q25.2 on chromosome 10. The present invention is exemplified by DNA cloned from the q24-q26 region of the mardel (10) chromosome as well

as the corresponding region on normal human chromosome 10. These DNA molecules contain a functional neocentromere. The present invention extends, however, to any neocentromere or any chromosome in mammalian and non-mammalian animals as well as plants, yeasts and fungi.

For convenience, the DNA clones from the mardel (10) chromosome as well as from normal human chromosome 10 are summarised in Figure 11. The neocentromere located at or around 10q25 is located on a clone designated the "HC-contig". DNA clones from mardel (10) are referred to as "E8" or the "NC-contig" which extends from the long arm (q') of mardel (10) towards the short arm (p'). Clone F2 extends further p' from E8 (see Figure 11). It is emphasised, however, that the present invention extends to any neocentromere on any human chromosome as well as neocentromeres on other mammalian and non-mammalian chromosomes including chromosomes from plants, insects, reptiles, yeast and fungi.

The present invention further contemplates a nucleic acid molecule or its chemical equivalent having a tertiary structure which defines a human neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof or its mammalian or non-mammalian homologue.

Even more particularly, the present invention is directed to an isolated nucleic acid molecule having a sequence of nucleotides or their chemical equivalents which directs a conformation defining a human neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof or its mammalian or non-mammalian homologue wherein the centromere associates with centromere binding proteins (CENP) -A and CENP-C or antibodies thereto.

Reference herein to "latent" in relation to a centromere includes reference to a centromere not normally functional but nevertheless activatable under certain conditions. A latent centromere may also be considered as a neocentromere provided it has no substantial α-satellite DNA repeat sequences.

The size of the neocentromere in accordance with the present invention may range from about

50 bp to about 1500 kbp, from about 70 bp to about 1000 kbp, from about 75bp to about 800 kpb, from about 80 bp to about 500 kbp, from about 85 bp to about 200 kbp, from about 90 bp to about 100 kbp, from about 100 bp to about 1 kbp, about 120 bp to about 500 bp, about 180 bp to about 300 bp. In one particular embodiment, the centromere is approximately 60-100 kbp. In another embodiment, the centromere is about 80 kbp.

The nucleic acid molecule encompassing the HC-contig for human chromosome 10 of the present invention set forth in Figure 6 (SEQ ID NO: 3). The nucleic acid molecule encompassing the NC-contig (part of E8) from mardel (10) is set forth in Figure 16A (SEQ ID NO: 4). The nucleic acid molecule encompassing F2 of mardel (10) is set forth in Figure 16B as separate contigs (SEQ ID NOs: 5-29). The nucleic acid molecules have a tertiary structure and the neocentromere is a conformation of nucleotides within this tertiary structure. Accordingly, the neocentromere is not defined by a linear sequence of nucleotides although this linear sequence directs the conformation which in turn defines the neocentromere. Although this aspect of the present invention is exemplified using the nucleotide sequence set forth in Figure 6, 16A and 16B, the subject invention extends to any sequence directing a conformation defining a centromere and hybridising to the sequence set forth in one or more of Figures 6, 16A and/or 16B under low stringency conditions at 42°C and/or which comprises a nucleotide sequence having at least about 40% nucleotide similarity to one or more sequences set forth in Figures 6, 16A and/or 16B. Preferably, the percentage similarity is at least about 50%, more preferably at least about 60%, still more preferably at least about 70%, even more preferably at least about 80-90% or above such as 95%, 97%, 98% and 99%.

Another embodiment of the present invention is directed to YAC 3 and YAC 5 encompassing the HC contig and flanking sequence as well as nucleotide sequences related to YAC 3 and/or YAC 5 at the homology, similarity or hybridization levels.

Reference herein to a low stringency at 42°C includes and encompasses from at least about 1% v/v to at least about 15% v/v formamide and from at least about 1M to at least about 2M salt for hybridisation, and at least about 1M to at least about 2M salt for washing conditions. Alternative stringency conditions may be applied where necessary, such as medium stringency, which

includes and encompasses from at least about 16% v/v to at least about 30% v/v formamide and from at least about 0.5M to at least about 0.9M salt for hybridisation, and at least about 0.5M to at least about 0.9M salt for washing conditions, or high stringency, which includes and encompasses from at least about 31% v/v to at least about 50% v/v formamide and from at least about 0.01M to at least about 0.15M salt for hybridisation, and at least about 0.01M to at least about 0.15M salt for washing conditions. These stringency conditions may be altered dependent on the source of DNA and other factors.

The term "similarity" as used herein includes exact identity between compared sequences at the nucleotide level. Where there is non-identity at the nucleotide level, "similarity" includes differences between sequences which nevertheless result in conformation defining a functional neocentromere.

The nucleic acid molecule of the present invention may comprise a naturally occurring nucleotide sequence from a healthy human subject or may comprise the nucleotide sequence from a human subject exhibiting one or more chromosomal-dependent conditions such as a subject carrying mardel 10 chromosome or a chromosome conferring an equivalent or similar condition or may carry one or more nucleotide substitutions, deletions and/or additions relative to the naturally or non-naturally occurring sequence. Such modifications are referred to herein as "derivatives" and include mutants, fragments, parts, homologues and analogues of the naturally occurring nucleotide sequence. Preferably, the derivatives of the present invention still define a functional neocentromere.

Reference herein to a "neocentromere" includes reference to a functional neocentromere or a functional derivative thereof meaning that it is capable of facilitating sister chromatid cohesion and chromosomal segregation during mitotic cell divisions and/or is capable of associating with CENP-A and/or CENP-C and/or is capable of interacting with anti-CENP-A antibodies or anti-CENP-C antibodies. Generally, and preferably, the neocentromere is incapable of interacting with CENP-B or anti-CEP-B antibodies. Alternatively, the neocentromere may be a latent centromere capable of activation by epigenetic mechanisms. The neocentromere may also be a hybrid of other human, mammalian, plant or yeast neocentromeres. Synthetic neocentromeres

provided by, for example, polymeric techniques to arrive at the correct confromation are also contemplated by the present invention. All such forms and definitions of neocentromere are encompassed by use of this term.

Another aspect of the present invention provides an isolated nucleic acid molecule or chemical equivalent having the following characteristics:

(i) comprises a nucleotide sequence or chemical equivalent directing a conformation which defines a neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof or;

(ii) comprises a nucleotide sequence or chemical equivalent substantially as set forth in one or more of Figures 6, 16A and/ or 16B or having at least about 40% similarity thereto or capable of hybridising thereto under low stringency conditions at 42°C; and

(iii) comprises a neocentromere capable of associating with CENP-A or CENP-C or antibodies thereto.

Preferably, the neocentromere is incapable of interacting with CENP-B or antibodies thereto.

In a particularly preferred embodiment, the centromere corresponds to a human genomic region which maps between q24 and q26 on chromosome 10, and in particular q25 on chromosome 10.

The nucleic acid molecule or its chemical equivalent of the present invention defining a conformational neocentromere or functional derivative thereof or latent, synthetic or hybrid form thereof is useful *inter alia* for the generation of artificial chromosomes such as human artificial chromosomes (HACs), mammalian artificial chromosomes (MACs), yeast artificial chromosomes (YACs) and plant artificial chromosomes (PLACs). HACs are particularly useful since they are capable of accommodating large amounts of DNA and are capable of propagation in human cells. The HACs are non-viral in origin and, hence, are more suitable for gene therapy by, for example, introducing therapeutic genes. Furthermore, the HACs remain extra-chromosomal and, hence, have no insertional/substitutional mutagenic potential. The essence of a HAC is the presence of a neocentromere or latent, synthetic or hybrid form thereof which enables stable segregation during cell division. The HAC also remains extra-chromosomal and, hence, is more suitable for

gene therapy. Reference to "extra-chromosomal" means that it does not integrate into the main chromosome and, in effect, is episomal.

Accordingly, the present invention provides a genetic construct comprising an origin of replication for a eukaryotic cell and a nucleic acid molecule encompassing a eukaryotic neocentromere or a functional derivative thereof or a latent, synthetic, hybrid form thereof or its mammalian or non-mammalian homologue flanked by telomeric nucleotide sequences functional in the cell in which the genetic construct is to replicate and wherein said genetic construct when introduced into a cell is a replicating, extra-chromosomal element which segregates with cell division.

More particularly, the present invention further contemplates a genetic construct in the form of an artificial chromosome comprising an origin of replication for a mammalian, human, plant or yeast cell and a nucleic acid molecule encompassing a human neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof or its mammalian or non-mammalian homologue flanked by telomeric nucleotide sequences functional in the cell in which the artificial chromosome is to replicate.

Another embodiment provides a genetic construct in the form of an artificial chromosome comprising an origin of replication for a mammalian, human, plant or yeast cell and a nucleic acid molecule having a tertiary structure which defines a human neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof or its mammalian homologue flanked by telomeric sequences functional in the cell in which the artificial chromosome is to replicate.

Yet another embodiment is directed to a genetic construct in the form of an artificial chromosome comprising an origin of replication for a mammalian, human, plant or yeast cell and a nucleic acid molecule having a sequence of nucleotides which directs a conformation defining a human neocentromere wherein the centromere associates with CENP-A and/or CENP-C or antibodies thereto and does not contain substantial α-satellite DNA repeat sequences, said nucleic acid molecule flanked by telomeric nucleotide sequences functional in the cell which the

artificial chromosome is to replicate.

Still yet another aspect of the present invention relates to a genetic construct in the form of an artificial chromosome comprising an origin of replication for a mammalian, human, plant or yeast cell and a nucleic acid molecule comprising a sequence of nucleotides which:

(i) directs a conformation which defines a neocentromere or a functional form thereof or a latent, synthetic or hybrid form thereof;

(ii) comprises a nucleotide sequence substantially as set forth in one or more of Figures 6, 16A and/or 16B or having at least about 40% similarity to the nucleotide sequences set forth in Figures 6, 16A and/or 16B or is capable of hybridising to one or more of these sequences under low stringency conditions at 42°C;

wherein the neocentromere is capable of associating with CENP-A and/or CENP-C or antibodies thereto and wherein said nucleic acid molecule is flanked by telomeric nucleotide sequences functional in the cell in which the artificial chromosome replicates.

In a preferred embodiment, the genetic construct is a HAC and comprises human telomeric sequences. In a particularly preferred embodiment, the HAC further comprises yeast artificial chromosome (YAC) arms and, hence, becomes a HAC/YAC shuttle vector capable of propagation in human and yeast cells. Preferably, the HAC/YAC contains a unique enzyme site between yeast telomeric sequences and human telomeric sequences such that upon contact with the particular enzyme, the yeast telomeric sequences are removed leaving the human telomeric sequences. Preferably, the unique enzyme site is a yeast specific enzyme site such as I-SceI.

According to this embodiment, there is provided a genetic construct defining a HAC/YAC comprising an origin of replication and a nucleic acid molecule encompassing a human neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof or a mammalian or non-mammalian homologue thereof, said nucleic acid molecule flanked by human telomeric sequences which are in turn flanked by yeast telomeric sequences wherein a unique enzyme site is located between the human and yeast telomeric nucleotide sequences such that upon contact with the enzyme, the yeast telomeric sequences are removed and the human telomeric sequences are exposed.

More particularly, the present invention is directed to a genetic construct defining a HAC/YAC comprising an origin of replication and a nucleic acid molecule encompassing a human centromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof or a mammalian or non-mammalian homologue thereof wherein the neocentromere associates with CENP-A and/or -C or antibodies thereto and does not contain substantial α-satellite DNA sequences wherein said nucleic acid molecule is flanked by human telomeric sequences which are in turn flanked by yeast telomeric sequences wherein a unique enzyme site is located between the human and yeast telomeric nucleotide sequences such that upon contact with said enzyme, the yeast telomeric sequences are removed and the human telomeric sequences are exposed.

Even more particularly, the present invention is directed to a genetic construct in the form of a HAC/YAC comprising an origin of replication and a sequence of nucleotides which directs a conformation defining a human neocentromere or a functional derivative thereof or a latent, synthetic or hybrid form thereof or a mammalian or non-mammalian homologue thereof wherein said neocentromere is capable of associating with CENP-A and/or CENP-C or antibodies thereto, said sequence of nucleotides flanked by human telomeric sequences which are in turn flanked by yeast telomeric sequences wherein a unique enzyme site is located between the human and yeast telomeric nucleotide sequences such that upon contact with said enzyme, the yeast telomeric sequences are removed and the human telomeric sequences are exposed.

Preferably, the length of the nucleotide sequence is between about 30 kpb and 1500 kpb, and more preferably between 60 kbp and 1000 kpb.

In a particularly preferred embodiment, the unique enzyme site is a yeast specific enzyme site such as I-SceI.

The present invention extends to yeast cells and human cells carrying the genetic constructs of the present invention and to proteins produced therefrom.

The genetic constructs may also comprise marker genes and other unique restriction sites to facilitate insertion of adventitious DNA. Accordingly, the genetic constructs of the present

invention may further comprise adventitious or heterologous DNA encoding a product of interest. Preferred products of interest include pharmaceutically useful genes such as genes encoding cytokines, receptors, growth regulators and the like. Endogenous genes may also be replaced by wild-type genes or modified genes.

The adventitious or heterologous DNA may also encode a molecule not synthesised in a sufficient amount in a particular subject and hence the increased copy number permits greater amounts of the molecule being synthesised.

Accordingly, the present invention contemplates a genetic construct comprising an origin of replication and a first nucleic acid molecule defining a human neocentromere or a functional derivative thereof or latent, synthetic or hybrid form thereof or a mammalian or non-mammalian homologue, a second nucleic acid molecule encoding a peptide, polypeptide or protein, wherein said first and second nucleic acid molecules are flanked by a first set of human telomeric sequences which are in turn flanked by a second set of yeast telomeric sequences wherein there are unique enzyme sites between the human and yeast telomeric sequences such that upon contact with said enzyme, the yeast telomeric sequences are cleaved off to expose the human telomeric sequences.

Reference herein to segregate preferably means mitotically stable segregation. Conveniently, stable segregation may be determined as the presence of an artificial chromosome in 40-60% of daughter cells after 4-6 months of continuous passage.

The present invention extends to other artificial chromosome analogues to the HACs and HAC/YACs described above such as MACs and PLACs.

Another aspect of the present invention relates to peptides, polypeptides and proteins which bind, interact or otherwise associate with the human neocentromere of the present invention or its mammalian and non-mammalian homologue. Preferably, the molecules are proteins, referred to as primary (1°) proteins. The 1° proteins bind to the neocentromere and secondary (2°) proteins bind to the 1° proteins before or after association with the neocentromere. The identification

of the human neocentromere in accordance with the present invention provides a mechanism for assaying 1° proteins and 2° proteins which may be important for screening chromosomes in, for example, genetic disorders. This is particularly the use in Down's Syndrome which results from defective chromosome segregation.

The 1° proteins are readily detected by, for example, a gel shift assay. The nucleic acid molecule of the present invention defining the human neocentromere is digested, labelled and contacted with nuclear extract putatively containing the 1° proteins and resolved on a gel. When a 1° protein binds to a fragment carrying a binding portion of the neocentromere, the DNA fragment migrates in the gel at a slower rate due to the bound protein.

The present invention extends to purified 1° proteins capable of association with the subject centromere and to genetic sequences encoding same and to antibodies thereto.

The neocentromeres of the present invention are readily identified and characterised using, for example, human fibrosarcoma cell lines. For example, DNA suspect of carrying a neocentromere, is introduced into fibrosarcoma cells in a linear form, generally together with a telomeric sequence. The cells are then screened for the presence of replicating, extra-chromosomal and segregating elements, referred to as mini chromosomes.

The present invention further encompasses eukaryotic cells carrying replicating, extrachromosomal and segregation nucleic acid molecules. Preferably the eukaryotic cells are mammalian cells and most preferably human cells. The nucleic acid molecules according to this aspect of the present invention are preferably as herein described. Particularly preferred cells are HT-38, HT-47, HT-54, HT-190, HT-191, BAC/E8-1, and BAC/F2-14.

The present invention is further described by the following non-limiting Figures and Examples.

EXAMPLE 1

YAC and Cosmid Probes for FISH

YACs carrying specific STSs were identified (Moir *et al.*, 1994) by PCR-based screening of YAC libraries prepared in pYAC4 vector at the Center for Genetics in Medicine at Washington University (Brownstein *et al.*, 1989) and at the CEPH (Albertsen *et al.*, 1990). Cosmid DNA inserts (35-40 kb) were ligated to SuperCos I vector (Stratagene) and packaged with Gigapack III Gold extract (Stratagene) according to the manufacturer's instructions. YAC probes were prepared by Alu-PCR of total yeast genomic DNA using primers 5'-GGATTACAGG(C/T)(A/G)TGAGCCA-3' [SEQ ID NO:1] and 5'-(A/G)CCA(C/T)TGCACTGCAGCCTG-3' [SEQ ID NO:2] according to published method (Archidiacono *et al.*, 1994). For probe labelling, 1 µg of the YAC PCR products or whole cosmid DNA isolated by CsCl centrifugation or Qiagen column was used. The DNA was labelled with Biotin-16-dUTP (Boehringer Mannheim) using a NICK translation kit (Boehinger Mannheim). A probe mix of 6-10 µg/ml of biotinylated probe DNA, 300 µg/ml of COT-1 DNA (Boehringer Mannheim), 500 µg/ml of carrier salmon sperm DNA and, where indicated, 10 µg/ml of biotinylated 10pC38 tag DNA was ethanol precipitated, resuspended in a hybridization mix of 50% v/v formamide in 2 x SSC and 10% w/v dextran sulphate, denatured at 95°C for 5 min, preannealed for 30-60 min at 37°C to suppress repetitive sequences, before adding to slides. FISH of α-satellite and satellite III probes was performed under low stringency as previously described (Voullaire *et al.*, 1993).

EXAMPLE 2

Somatic Cell Hybrids and Other Cell Lines

Skin fibroblasts and transformed lymphoblast cell lines were established from patient BE (Voullaire *et al.*, 1993) and from his normal parents. The presence of the mardel 10 chromosome in the patient cell lines was confirmed by FISH. In addition to these cell lines, two somatic cell hybrids were produced by fusing cultured fibroblast cells derived from patient BE with the Chinese hamster ovary cell line CHO-K1 using polyethylene glycol. Hybrid cells were selected in a proline-free medium for the glutamic oxaloacetic transaminase-1 (GOT-1) gene located in

10q24-q25 region. One of the hybrid cell lines, designated BE2C1-18-1f, was shown to contain the normal chromosome 10 but not the marker chromosome, while another hybrid cell line, designated BE2C1-18-5F, contained the marker chromosome but not the normal chromosome 10 of patient BE. The presence or absence of these chromosomes was established by karyotyping and ANTI-CEN/FISH probing. In addition, PCR analysis of an STS (sequence tagged site) marker, AFM259xg5, which resided on YAC-3, confirmed the status of these chromosomes in the hybrids and excluded the presence of submicroscopic fragments of the marker centromere region within the genome of BE2C1-18-1f, or the presence of the corresponding region of normal chromosome 10 within the genome of BE2C1-18-5f. Use of this STS marker also demonstrated that the mardel 10 chromosome has originated from the patient's father.

EXAMPLE 3

Antisera

Antiserum CREST #6 was from a patient with calcinosis, Raynaud's phenomenon, esophageal dysmotility, sclerodactyly and telangiectasia (a constellation of symptoms commonly referred to as "CREST"; Moroi *et al.*, 1981; Fritzler and Kinsella, 1980; Brenner *et al.*, 1981). Western blot analysis of this antiserum indicated that the primary antigens detected were human CENP-A and CENP-B. A specific anti-CENP-C polyclonal antibody, designated Am-C1, was produced by the inventors by expressing a partial mouse CENP-C polypeptide (amino acid #41 to 345) as a GST-fusion product in *E. coli*, followed by gel purification of the product and its use as an antigen for antibody production in rabbit.

EXAMPLE 4

Preparation of Standard Metaphase Chromosomes for FISH analysis

Actively replicating transformed lymphoblasts were incubated at 37°C for 17 h in the presence of 0.1M final concentration of thymidine before they were centrifuged at 2000 rpm for 10 min, washed with pre-warmed RPMI, and incubated for a further 5-6 h. 15 min before harvesting, colcemid (10μg/ml) was added. Cells were harvested according to standard cytogenetic techniques using 0.075M KCl hypotonic solution for 15 min at 37°C, followed by three fixative

washes in ice cold methanol/acetic acid 3:1, dropped onto clean glass slides, and stored dessicated at -20°C until required.

EXAMPLE 5

Preparation of Mechanically Stretched Chromosomes for ANTI-CEN/FISH Mapping

METHOD - I

This is an adaptation of the method described by Page *et al.* (1995). Colcemid (10μg/ml) was added to actively dividing transformed lymphoblasts for 2-3 h, before the cells were centrifuged at 1500 rpm for 10 min, washed in PBS, and resuspended in 0.075M KCl hypotonic solution for 10 min at RT at a concentration of approximately 5 x 10^4 cells/ml; the use of fewer cells here gave better stretching of the chromosomes. 200-300µl of this suspension were then cytocentrifuged onto clean microscope slides using a Cytospin 2 (Shandon) at 1000 rpm for 5 min at high acceleration. The slides were immediately removed, placed flat in a shallow dish and very gently flooded with KCM (Potassium Chromosome Medium:120 mM KCl, 20mM Nacl, 10mM Tris-HCl, 0.5mM Na_2EDTA, 0.1% v/v Triton X-100) (Jeppesen *et al.*, 1992). After 10 min at RT, immunofluorescence was performed without fixation (Earnshaw and Migeon, 1985; Earnshaw *et al.*, 1989; Jeppesen *et al.*, 1992; Jeppesen and Turner, 1993). KCM buffer was gently aspirated and 50µl of CREST#6 serum [diluted 1:50 in 1 x TEEN (1mM Triethanolamine HCl, 0.2 mM Na_2EDTA, 25 mM NaCl), 0.1% v/v Triton X-100, 0.1% w/v BSA] was added to the cell area of the slide and covered with a parafilm coverslip. The slides were incubated for 30 min at 37°C, then washed very gently by flooding in 1 x KB$^-$ [10 mM Tris-HCl (pH7.7), 0.15M NaCl, 0.1% w/v BSA), three rinses of 3 min each at RT. The primary antibody was detected with Texas Red-conjugated Affini-pure Rabbit anti-Human IgG (H&L) (Jackson Laboratories) diluted 1:50 in 1 x KB$^-$. 50 µl was added to each slide, covered with a parafilm coverslip, and incubated for 30 min at 37°C. The slides were again gently washed by flooding in 1 x KB$^-$ for 2 min at RT, before they were fixed by flooding in 10% v/v formalin in KCM for 10 min at RT, followed by three rinses of 3 min each in distilled water. If FISH was not performed the slides were rinsed in PBS and mounted in DAP1 (0.25 µg/ml) in DABCO antifade mountant. [In experiments where CREST#6 and Am-C1 antisera were simultaneously used to label the centromere (Figs. 2B and C), the above procedure was

followed except for the addition of Am-C1 diluted 1:100 together with CREST#6, and the Am-C1 antibody was detected using 1:100 diluted Donkey anti-Rabbit DTAF (Jackson Laboratories)].

If FISH was to be performed on the slides, they were then given a second fix in 3:1 methanol/acetic acid for 15 min at RT. The slides were air dried for at least 5 min and either processed for FISH or stored at -20°C for up to several days before continuing. For FISH, the slides were dehydrated at RT in 70%, 90%, 100% v/v ethanol (2 min each) and air dried. Chromosomal DNA was denaturated in deionised 70% v/v formamide/2xSSC, pH 7.0 at 82°C for 8 min followed by immediate dehydration in 70%, 90% and 100% v/v ethanol at -20°C for 2 min each, then air dried for at least 10 min. (This high temperature of denaturation was critical to obtain maximum FISH signals). An amount of 15 µl of the prepared probe was added to each slide, covered with a 22mm^2 coverslip, and sealed with rubber cement. Slides were hybridized overnight in a humid chamber at 37°C, then rinsed in 2 x SSC at RT, followed by 3 washes of 0.1 x SSC at 60°C for 5 min each, rinsed again in 2 x SSC, and immersed in a blocking agent of 5% non fat milk in 4 x SSC for 10 min at RT. Probe hybridization was detected by incubation with FITC-conjugated avidin at 37°C for 30 min, followed by three washes of 5 min each at RT in wash buffer (4 x SSC, 0.05% v/v Tween-20). Signals were amplified by incubating with goat anti-avidin D antibodies for 30 min at 37°C, followed by three washes of 5 min each at RT in wash buffer, then with another layer of avidin-FITC for 30 min at 37°C, before the slides were washed in wash buffer, rinsed in PBS, and counter-stained with DAP1 (0.25 µg/ml) in DABCO mountant.

METHOD - II

The following method was modified from that of Haaf and Ward, (1994). Actively dividing lymphoblast cells were treated with 10µg/ml colcemid for 2-3h, washed in PBS and resuspended in a hypotonic solution consisting of 10mM Hepes (pH7.3), 30mM glycerol, 1.0mM $CaCl_2$ and 0.8mM $MgCl_2$, at a cell density of approx. 2.5 x 10^2/ml. After 10 min of hypotonic treatment at RT, 300 µl were cytocentrifuged (Shandon - Cytospin 2) onto glass slides at 800 rpm for 4 min. The slides were immediately removed from the centrifuge, dried

for 15 sec, fixed in methanol at -20°C for 20-30 min, rinsed in acetone at -20°C for a few sec, then washed in 3 rinses of PBS at RT. Immunofluorescence staining was done using CREST#6 at a dilution of 1:50 in PBS. After incubation at 37°C for 30 min, the slides were washed three times in PBS for 2 min each. This primary antibody was then detected by a further incubation for 30 min at 37°C with Texas Red-conjugated Rabbit anti-Human IgG diluted at 1:50 in PBS. The slides were fixed in 10% v/v formalin in KCM for 10 min at RT, then washed in 3 rinses of distilled water and drained. Before FISH was performed, slides were fixed in methanol/acetic acid 3:1 for 15 min at RT and air dried. Chromosomal DNA was denatured in 70% v/v deionised formamide (pH7.0) in 2 x SSC at 82°C for 4-6 min. After dehydration in an ice cold ethanol series the slides were air dried, and used for FISH as described for Method I. Slides could be stored covered in foil at RT after methanol/acetic acid fix for up to several weeks before FISH.

Both methods I and II were used to obtain the results shown in Figs. 2B, 2C, 3 and 4B.

EXAMPLE 6

Image Analysis

Hybridization signals for YAC mapping on standard metaphase preparations utilized a normal fluorescence microscope. Images for the ANTI-CEN/FISH experiments were analyzed on a Zeiss Axiolab fluorescence microscope equipped with a 100x objective and a cooled CCD camera (Photometrics Image Point) controlled by a Power Mac computer. Gray scale images were captured separately using a LUDL filter wheel and controller for Texas Red, FITC and DAPI. These images were pseudocoloured and merged using IPlab Spectrum software from Signal Analytics Corporation. A number of difficulties were commonly associated with the ANTI-CEN/FISH technique: (a) the deliberate "stretching" of the chromosomes, whilst increasing the resolution of mapping, sometimes caused serious distortion to the chromosomes, often making them quite dysmorphic; (b) FISH treatment following the ANTI-CEN-labelling often significantly reduced the ANTI-CEN signals; (c) more highly stretched chromosomes (which would potentially give better mapping resolution) generally gave weaker ANTI-CEN signals; and (d) the ANTI-CEN signal on the mardel 10 centromere was usually weaker than

those of the other human chromosomes. Thus, a cell would only be considered informative and used for scoring if both the p'- and q'-arms of the mardel 10 chromosome were discernible and separated by a discrete ANTI-CEN signal. In addition, FISH signals for both the test probe and the 10pC38 cosmid tag (used to identify the q'-arm of, and thus orientate, the marker chromosome) must be clearly present. Using these criteria, the overall frequency of informative cells was found to be approximately 1 in every 20-30 metaphases analyzed.

EXAMPLE 7

Restriction Analysis of Patient DNA

High-molecular weight genomic DNA was extracted from cultured fibroblast cell lines of patient BE and those of his parents and digested with different enzymes to generate restriction fragments ranging from <1kb up to ~1 Mb. The digested DNA was resolved either on a standard agarose gel or by pulsed-field gel electrophoresis (PFGE) using a Bio-Rad CHEF-XA Mapper. For filter hybridization, 50-100 ng of whole cosmid or PAC DNA was labelled by random priming. The labelled probe was then added to 2 ml of hybridization buffer (0.5M Na_2HPO_4, 7% w/v SDS, 1% w/v BSA, 1mM EDTA, pH. 7.0) containing 500 µg of human placental DNA (Sigma). The mixture was boiled for 5 min, then placed in a 65°C water bath for preannealing of repetitive DNA for 90 min. The preannealed probe mix was then added to prehybridizing filters and hybridized overnight at 65°C. Post-hybridization washes were at a final stringency of 0.1 x SSC, 0.1% w/v SDS at 68°C.

EXAMPLE 8

Identification of a YAC region spanning the marker centromere

The initial search for DNA sequences spanning the centromere of the mardel 10 chromosome was based on fluorescence in situ hybridization (FISH) of existing cosmid and YAC clones (Moir *et al.*, 1994; Zheng *et al.*, 1994) that have been mapped to the q24 - q26 region of the normal human chromosome 10 where the new marker centromere was formed (Voullaire *et al.*, 1993) (Fig. 1A). This search led to the identification of a 4 megabase YAC contig (designated #082) that spanned the marker centromere region (Fig 1B). Fig. 1C graphically presents the

FISH mapping results with selected YACs from this contig. As can be seen, two of the YACs (YACs-1 and YAC-2) mapped to the q'-side of the marker centromere, whereas the remaining YACs mapped to the p'-side of the centromere. The low signal level observed for YAC-3 was due to a large proportion of this probe hybridising directly on the centromere itself. These results, therefore, provided evidence that YAC contig #082 spanned the marker centromere, and that the centromere region was likely to be within YAC-3, where the "cross-over" between the q' and p' signals occurred.

EXAMPLE 9

Development of Improved ANTI-CEN/FISH Methods for the Simultaneous Detection of Marker Centromere and Single-copy Cosmid DNA Probes

Although normal fluorescence microscopy and FISH analysis of standard metaphase chromosomes were adequate for the initial identification of the YAC contig spanning the marker centromere, methods with significantly higher sensitivity and resolution were needed to allow further walking into the marker centromere DNA. Three requirements have to be satisfied by these methods: (a) the metaphase chromosomes have to be extended to offer much greater mapping resolution, (b) the centromeres have to be more precisely defined than that offered by a cytogenetic constriction, and (c) the methods should allow simultaneous visualization of both the centromere antibody and FISH signal. Two published methods were explored (designated here as ANTI-CEN/FISH methods) based on extending metaphase chromosomes by mechanical stretching and labelling of the neocentromere by autoimmune antibodies (Haaf and Ward, 1994; Page *et al.*, 1995). Since these methods were originally established for the labelling of normal centromeres and for FISH analysis of highly repeated DNA, they were modified (see Example 4) to allow detection of the generally reduced ANTI-CEN signal of the subject marker neocentromere and the lower FISH signals resulting from the use of single-copy cosmid DNA probes.

With the improved detection methods, the status of α-satellite and satellite III DNA on the marker neocentromere was reassessed, since this was previously determined using standard

microscopy and FISH (Voullaire *et al.*, 1993). Fig. 2A shows the result of antibody labelling using CREST#6 and FISH using α-satellite DNA, and indicated the absence of detectable signal on the marker centromere. The same result was obtained when the experiments were repeated without ANTI-CEN-labelling, ruling out the possibility that the anti-centromere antibody might have obscured any weak FISH signals. Similar results were obtained with satellite III DNA. Since in separate reconstruction experiments, it was possible to demonstrate the sensitivity of the procedure in detecting a single-copy DNA probe of less than 1.5 kb, and making the reasonable assumption that the low-stringency hybridization conditions used for the α-satellite and satellite III DNA which, by virtue of the use of >100-fold excess of probes and the strong hybridisation of these probes to all the other centromeres, would have allowed the detection of any related sequences, it can be concluded that these satellite are absent.

SUBSTITUTE SHEET (RULE 26)

EXAMPLE 10

Co-localization of CENP-C and CENP-A on the Marker neocentromere

To test if CENP-C is present on the marker centromere, a specific rabbit polyclonal antibody was prepared against a recombinant product of mouse CENP-C. This antibody, designated Am-C1, reacted strongly with the centromeres of rodent and human chromosomes. Fig. 2B shows results for the labelling of stretched human metaphase chromosomes using this antibody simultaneously with the CREST#6 autoimmune antibody. As can be seen, irrespective of the degree of chromosome stretching, the signals for the two antibodies coincided fully on all the centromeres. The localization of these two antibodies on the marker chromosome was further determined by employing the 10pC38 cosmid tag in an ANTI-CEN/FISH experiment to identify the marker chromosome. The results indicated that both the antibody signals were clearly present and again coincided completely on the marker centromere (Fig. 2C, a-e). Although CREST #6 was known to bind CENP-A and CENP-B, indirect evidence suggests that binding to the marker centromere presumably occurred *via* CENP-A since the presence of the marker centromere was previously demonstrated not to bind CENP-B (Voullaire *et al.*, 1993). The above results, therefore, established the localization of CENP-C, and probably CENP-A, on the marker centromere.

EXAMPLE 11

Localization of the anti-centromere antibody-binding domain

For further walking into the marker centromere region, cosmid libraries were prepared from total yeast genomic DNA containing YACs-2, -3, -4, -6, -7, -13, and -17. Cosmid clones containing human DNA inserts were isolated by hybridization with human COT-1 DNA using low stringency. All resulting cosmids were screened by standard FISH to confirm their localization to the expected marker centromere and normal chromosome 10 regions, and to eliminate clones that might have originated from other genomic sites due to chimeric YACs. Positive clones were then analyzed further with the ANTI-CEN/FISH methods, using CREST#6 to label the centromere. Fig. 3*a* (I and II) show examples of cosmid signals that mapped to the q'- and p'-side, respectively, of the marker centromere in the ANTI-CEN/FISH experiments. The cosmid tag (clone 10pC38) was used in these experiments to define the q' arm of the marker chromosome. For cosmid walking, we concentrated on clones derived from YAC-3 since FISH mapping of YAC contig #082 indicated that the marker centromere region was likely to be within this YAC. Fig. 4*a* shows a restriction map of the region covered by this and surrounding YACs and compares this map with a genomic map derived from patient BE. The relative positions of a series of cosmid clones (including five independent PACs) were also determined and placed on the YAC map. Fig. 4*b* presents the ANTI-CEN/FISH results obtained with a number of the cosmid clones and one of the PAC clones. Clones Y3C64, Y6C8, and Y3C94 localized preferentially to the q'-side, while Y13C1+C8 and Y17C6 localized preferentially to the p'-side of the marker centromere, suggesting that the nucleus of the antibody-binding domain is situated between these two cosmid clusters. Within this central region, a group of cosmid clones comprising the HC-contig (Fig. 4*a*) was found to map closely around the ANTI-CEN signal. Fig. 4*c* shows a restriction map for eight different overlapping clones from this HC-contig. The chromosomal positions of five of these overlapping clones were analyzed in detail using ANTI-CEN/FISH. Fig. 4*b* shows the cumulative results for more than 60 informative chromosomes for each of these five probes. The results indicated that Y7C14 mapped preferentially q'- of the antibody-binding domain, while the remaining four clones (Y4C45, Y6C10, Y6C21 and Y3C3) mapped preferentially to the p'-side. In addition, the results for PAC5 (a 75 kb-insert PAC clone that overlapped with the p'-end of PAC4 by approximately 5 kb; see Fig. 4*a*) provided further

evidence for the emergence of the HC-contig region onto the p'-arm. Based on these results, we conclude that the eight contiguous cosmid clones within the HC-contig shown in Fig. 4*c*, which together constitute ~80 kbp of DNA, have defined the nucleus of the antibody-binding domain of the marker centromere.

From the above ANTI-CEN/FISH results, it was difficult to determine if the sequences of the HC-contig and its surrounding DNA, both originally derived from a normal individual, were part of the marker centromere DNA, or whether these sequences simply flanked a transposed centromere DNA with an unrelated nucleotide composition. However, supporting evidence from the ANTI-CEN/FISH experiments suggested that the DNA of the HC-contig region appeared to be a part of the marker centromere. This came from the mapping of Y6C10 and Y6C21 onto superstretched chromosomes that were occasionally detected in the slide preparations. An example of such mapping is shown in Fig. 3*b* using Y6C21. As can be seen, whilst a significant portion of Y6C21 hybridized to the p'-side of the CREST signal on the highly extended chromosome, a substantial portion of the cosmid DNA also overlapped directly with the CREST signal. This suggests that at least part of the HC-contig region actually comprises the same DNA sequence as the marker centromere. This possibility was further investigated by detailed genomic mapping.

EXAMPLE 12

The Marker Centromere DNA has a Similar or Identical Sequence Organization as the HC-Contig

The genomic organization of the HC-contig region was compared with that of the corresponding DNA region of the mardel (10) chromosome. Three overlapping cosmids (Y7C14, Y6C10, and Y4C7, the latter being essentially the same as Y6C21; Fig. 4C) from the HC-contig were used as probes to analyze the restriction patterns of genomic DNA prepared from patient BE and those of his karyotypically normal parents. Fig. 5 shows examples of the band patterns obtained with Y6C10, while Table 1 summarizes the results for all the enzymes tested with Y7C14, Y6C10 and Y4C7. The detection of a single band on PFGE gels with a number of the enzymes indicated that the cosmid DNA sequences were unique within the

human genome (*SfiI*, *SalI*, *KspI*, *KpnI* and *BclI* in Fig. 5A; Table 1). The detection of a single on PFGE gels with a number of the enzymes (*ClaI* in Fig. 5A; Table 1) could be explained by differential methylation of different restriction sites found in this region (Nelson and McClelland, 1991); the reproducibility of these multiple band patterns ruled out incomplete digestion as a possible cause. The multiple bands detected with the more frequent cutting enzymes on a standard gel (Fig. 5B and Table 1) were a result of the presence of cleavage sites present within the probe DNA, since similarly digested cosmid DNA electrophoresed next to the genomic DNA yielded identical patterns for all the bands not containing cosmid vector sequences. In all, 37 enzymes were used to generate more than 160 different fragments for the three cosmid probes (Table 1). The results indicated that, except for a polymorphic fragment found in one of the parents, an identical banding pattern was present in the genomic DNA of patient BE and those of his parents. Furthermore, when the restriction patterns obtained for the genomic DNA of patient BE were compared with those of the smatic hybrid cell line BE2C1-18-5F, which contained the marker chromosome but not the normal chromosome 10, no detectable difference was seen between the two DNA preparations within the HC-contig region (Fig. 5C).

In addition to Y7C14, Y6C10 and Y4C7, a host of other probes from within or surrounding the HC-contig have been tested, each with an average of 12 different informative enzymes. These probes included PAC4 (which spanned the entire HC-contig region shown in Fig. 4C), Y3C64, Y3C109, Y6C6, Y6C8, Y3C94, PAC1, Y3C90, Y4C4, Y4C8, Y4C13, and Y3C33. The results again indicated identical restriction enzyme patterns between patient BE and normal DNA. Thus, through the analysis of a relatively large number of probes covering about 500 kb of YAC-3 around the HC-contig region, and the use of a high density of restriction enzymes that generated a range of fragments from <1 kb to ~1 Mb, it was evident that the marker centromere DNA and a substantial stretch of its adjoining regions showed no detectable difference against the corresponding genomic region of the normal chromosome 10.

Since a potential limitation of the above Southern blot analyses was that highly repeated sequences were not detected because of the preannealing step used in the hybridisation procedure, a different approach was employed to compare the DNA of the marker chromosome

and that of the normal chromosome 10. In this approach, oligonucleotide primers from different regions of the HC-contig were used to prepare a series of PCR fragments from the BE2C1-18-5F and BE2C1-18-1F hybrid cell lines. Electrophoretic comparison of such fragments, which randomly covered approximately 40 kb of the HC-contig, indicated no detectable difference between the two chromosomes and provided independent support for the results obtained in the Southern blot analyses. Thus, it can be concluded that the sequence organization of the marker centromere region is similar, if not identical, to that found in the HC-contig region of the normal chromosome 10.

EXAMPLE 13

Implications for Centromere Study and Mammalian Artificial Chromosome Construction

The mammalian centromere has been difficult to study due to the massive amount of repetitive DNA normally associated with it. By avoiding such repetitive DNA and analyzing the unusual centromere found in the present marker chromosome, the inventors have created a much more tractable system for centromere studies. The present analysis has already shed some light on the important question of DNA sequence versus conformational requirement of a centromere, and on the intriguing concepts of latent centromeres and epigenetic mechanisms. One urgent application of this DNA is to use it to identify the primary protein(s) which binds to the centromeric DNA. Another important application of the marker centromere DNA is in the construction of mammalian artificial chromosomes. Such artificial chromosomes offer a potentially powerful vehicle for the structural and functional analysis of chromosomes, for the genetic manipulation of plants and animals, and for the stable transmission of therapeutic genes in human gene therapy. The artificial chromosomes require a functional mammalian centromere, and the marker centromere DNA element of the present invention now provides a suitable centromere especially because of its relatively small size in the absence of α-satellite DNA and its cloning stability, as indicated by the cosmid, YAK and BAC clones of the HC-contig and NC-contig.

EXAMPLE 14

Sequence analysis

Figures 6, 16A and 16B show partial nucleotide sequences for the HC-contig (SEQ ID NO: 3) NC-contig [SEQ ID NO: 4] and F2 (BAC/F2-14) [SEQ ID NO: 5-29] regions, respectively.

EXAMPLE 15

Human Artificial Chromosome (HAC)

The following are examples of the different approaches being used in the inventors' laboratory for the production of a HAC:

Retrofitting of HC-contig DNA from normal chromosome 10

This procedure aims to produce HACs of 100 kb to >1Mb using the region of the normal chromosome 10 containing and surrounding the HC-contig DNA. The generation of a HAC by this approach will provide crucial proof that this normal DNA region can be reactivated to form a functional centromere.

A retrofitting procedure suitable for introducing human telomeres to both ends of any YAC prepared in the pYAC4 vector in the yeast host strain AB1380 has been previously described (Larin *et al.*, 1994; Taylor *et al.*, 1994, 1996). YACs (in particular YAC-3 and YAC-5) spanning the normal HC-contig region are used for retrofitting by plasmid constructs designed to recombine with their pYAC4 vector arms (Figure 7). The construct pLGTEL 1 is used to target the left arms of the YACs. This serves to add a LYS2 yeast selectable marker, gpt element for ultimate selection in mammalian and avian cell culture, and a human telomere. The right arm of the YACs are targeted by homologous recombination with pRANT 11 to produce a final construct where additional markers are introduced along with a second human telomere to cap the construct. Specifically, an ADE2 yeast marker is added and the URA3 gene of the YAC is disrupted, serving a useful role in negative selection of the construct. A neomycin (neo) resistance gene shown to function in mammalian and avian cells is also introduced. The finished

constructs are transfected into different cultured cell lines, including HT1080 (of human sarcoma origin) (Larin *et al.*, 1994; Rasheed *et al.*, 1974), DT40 (a recombination-proficient chicken cell line) (Dieken *et al.*, 1996), and BE2CI-18-5f (a human/hamster somatic hybrid cell line containing the mardel (10) chromosome but not the normal chromosome 10).

5

***In vitro* cloning of HC-region into YAC/HAC vectors**

The different vectors used for the cloning of the normal and mardel (10) centromeric DNA in the preparation of HACs are summarised in Table 2.

10

A number of different YAC cloning strategies are employed:

Conventional YAC cloning approach. Figures 8A-D show the different vectors used for cloning DNA as YACs by the conventional restriction/ligation methods. These YACs can then 15 be shuttled into mammalian cells and tested for HAC function.

ALU-ALU circular TAR cloning approach. Transformation-associated recombination (TAR) in the yeast *S. cerevisiae*, is a method for constructing linear and circular YACs from mammalian DNA (Larionov *et al.*, 1996a, 1996b). The recombination process is shown in Figure 9. Briefly, 20 the technique involves the use of a vector (pVC39-AAH2, Fig. 8E) lacking an autonomous replicating sequence (ARS) but containing a functional yeast centromere (e.g. CEN6) and selectable marker (e.g. HIS3), and two ALU DNA hooks to trap mammalian DNA by recombination at ALU sequences after co-transformation of linearized vector and high molecular weight DNA into yeast spheroplasts and followed by selection on medium lacking histidine. The 25 key to the process is that the mammalian DNA provides an ARS (11-bp sequence found frequently in mammalian DNA) which allows the HIS^+/CEN vector to replicate as a circular YAC. These YACs are very stable and range in size from 100 kb to greater than 600 kb (Larionov *et al.*, 1996b).

30 pVC39-AAH2 vector is used to clone DNA from hybrid BE2CI-18-5f to make YACs with an average insert of 250 kb. This TAR vector is further modified to create pAAH-TCNa (Fig. 8G)

so that it has the ability to shuttle between yeast and mammalian cells (as outlined in Figure 10), including the potential to expose human telomeres (TEL) at each end of a cloned fragment using a unique restriction site I-SceI.

Semi-specific and specific circular TAR. A modified circular TAR method utilising two specific 5'C and 3'C DNA hooks (300-700 bp in size) may be used to clone a specific human DNA at a frequency of 3/1000 HIS^+ transformants. The inventors prepared the vectors pVC39-ALU/C3-F2(+/-) and pTCN-TCS (Table 2) to perform semi-specific and specific TAR cloning, respectively.

The Semi-specific TAR methodology is a modification of a specific circular TAR strategy which permits the site directed isolation of target chromosomal DNA. Furthermore, in accordance with the present invention, the methodology described herein enables the site-specific cloning of target chromosomal DNA from total genomic DNA as a circular YAC at relatively high frequencies and without the need for the construction and extensive screening of complex libraries made from genomic DNA.

In a preferred embodiment of the present invention, the methodology employs a single specific DNA hook which flanks the mardel (10) chromosome and a less specific Alu-hook to trap the other side of the target DNA.

In initial experiments, a unique repeat DNA-free, 1.4kb *Eco*RI fragment (designated C3-F2) was identified from the p' side of the 80-kb HC-contig (Fig. 11A) (du Sart *et al.*, 1997). This fragment was subcloned into the centromere-based yeast circular TAR vector, pVC39-AAH2, by replacing the existing BLUR13 *Alu* (Larionov *et al.*, 1996b) to create the pVC39-ALU/C3-F2 constructs. As the specific orientation of the C3-F2 sequence on the chromosome was not known, the fragment was cloned in two different orientations, for which the (+) orientation (Fig. 11B) was expected to trap the genomic region to the left of C3-F2, while the (-) orientation was expected to trap the region to the right. Both constructs were used in yeast transformation.

As a source of genomic DNA containing the neo-centromere, a somatic hybrid cell line, BE2C1-

18-5f (du Sart *et al.*, 1997), containing the mardel 10 chromosome but not the normal human chromosome 10 was used. 5μg of high-molecular-weight DNA from this cell line and 1μg of pVC39-ALU/C3-F2(+) or pVC39-Alu/C3-F2(-) (linearized with *Sma*I to expose the 0.21-kb *Alu* and 1.4-kb C3-F2 hooks) were co-transformed into 10^9 (previously prepared and stored frozen) spheroplasts of S. *cerevisiae* YPH857 which carries a HIS3 gene deletion, (Sikorski and Hieter, 1989) and grown on SD, without HIS medium, (Larionov *et al.*, 1996a;b) to yield between 10 and 100 HIS^+ colonies. Control experiments in which YPH857 was transformed with vector alone did not produce any colonies, indicating that the C3-F2 fragment lacked ARS-like sequences. Twenty TAR experiments were performed and HIS^+ colonies were picked into 96-well trays containing YPD medium (supplemented with 50μg/ml ampicillin and 15μg/ml tetracycline), grown at 30°C with aeration for 24h and stored in 20% (v/v) glycerol at -70°C. Total yeast DNA was prepared in pools of 48 (Kwiatkowski jr *et al.*, 1990) and screened by PCR with the primers norm 5 and norm 7 (Table 3) which are located 30-kb q′ of C3-F2 (Fig. 11A). Two desired positive clones, designated 5f-52-E8 and 5f-38-F2, which contained the neo-centromere DNA derived from mardel 10 and mardel (10) and the DNA immediately p′ of the neocentromeric DNA, respectively, were identified. For subsequent studies, these clones were grown on SD without HIS medium and single colonies were re-isolated for characterization.

Initially, the sequence nature and sizes of the 5f-52-E8 and 5f-38-F2 insert DNA were determined. High-molecular-weight DNA was prepared in agarose blocks and digested with an enzyme (*Srf*I) that linearized with YAC (Fig. 11A). The linearized DNA, as well as uncut intact DNA, were resolved by pulsed-field gel electrophoresis (PFGE), transferred onto a nylon membrane and probed with radiolabelled PAC4, a P1-derived artificial chromosome clone containing a 120-kb insert that spans the entire HC-contig from normal chromosome 10, (du Sart *et al.*, 1997) following preannealing with human placental DNA to suppress repetitive DNA. The intact 5f-52-E8 and 5f-38-F2 remained trapped in the electrophoretic wells and the linearized DNA migrated into the gel and demonstrated a size of approximately 110 kbp and 80 kbp, suggesting insert sizes of about 105 kbp and 75 kbp, respectively (given that the vector size is 5.9 kb).

Despite the use of a genomic DNA source previously shown by sequence-tag-site (STS) analysis

to be free from normal chromosome 10 material, it is desirable to independently confirm the mardel (10) -origin of the 5f-52-E8 YAC clone. This was achieved using a set of primers (norm 17 and 18; Fig. 11A) that detected a variable-number-tandem repeat (VNTR) region within the HC-contig/neocentromere region. The results clearly indicated the presence of a 1.4-kb PCR product that was specific for the mardel (10) chromosome (Table 3).

PCR was used to further compare the 5f-52-E8 DNA with the previously cloned HC-contig sequence derived from normal chromosome 10. PCR products with sizes ranging between 0.2 and 15.9 kb were generated by standard PCR or with the Expand Long Template PCR system (Boehringer-Manneheim). Products greater than 1 kb were digested with frequent cutting enzymes, *Rsa*I and *Bsi*XI, and their fingerprints were compared by agarose gel electrophoresis. The results, shown in Table 3, indicated the absence of any detectable difference between the 5f-52-E8 DNA and those of the corresponding regions of the normal chromosome 10 (in somatic cell hybrid BE2C1-18-1f) and the neocentromere region of mardel (10) (in somatic cell hybrid BE2C1-18-5f). These results also demonstrated that the YAC 5f-52-E8 spanned at least 75 kb of the HC-contig region (Fig. 11C), consistent with the size determined by PFGE. Furthermore, the ability of all the internal primers to amplify DNA from 5f-52-E8 strongly suggested that the YAC was not chimeric. This result was confirmed by isolating DNA from four single-colony isolates of 5f-52-E8, digesting these with *Eco*RI and *Eco*RV, and probing with radiolabelled PAC4. The hybridization patterns obtained with these enzymes were consistent with those established in the previous study (du Sart *et al.*, 1997). Thus, this analysis, based on cloned DNA derived directly from mardel 10, has provided confirmation that the neocentromere DNA region is structurally identical to that of the corresponding HC-contig region of the normal chromosome 10 (du Sart *et al.*, 1997).

The circular YACs 5f-52-E8 and 5f-38-F2 were further retrofitted with the yeast-bacterial-mammalian cells shuttle vector BRV1 as previously described (Larionov *et al.*, 1997). The resulting BAC clones were designated BAC/E8-1 and BAC/F2-14, respectively (Fig. 11D).

The specific TAR strategy is outlined in Figure 12 and uses unique fragments from the HC-contig region, such as the ends of PAC4 (a 120 kb-insert PAC clone containing the HC-region)

to create the YAC/HAC shuttle vector pTCN-TCS. An example of a YAC/HAC construct containing the HC-contig region of normal chromosome 10 is shown in Fig 13.

Completed constructs are transfected into different cultured mammalian or chicken cells (see above) by lipofection using Transfectam or DOSPER.

In vivo "cloning" of HC-region into HAC vectors

This strategy employs a technique known as Telomere Associated Chromosomal Truncation (TACT) (Fig. 14). The technique is based on the principle that cloned mammalian telomeric DNA when reintroduced into a mammalian cell can seed the formation of a new telomere at an intrachromosomal location. If the introduced telomeric DNA is targeted to a known site through homologous recombination, integration at that location and subsequent truncation of distal sequences on the original chromomosomal arm can result (Brown *et al.*, 1994; Farr *et al.*, 1995). This technique is employed in our own study to truncate the mardel 10 chromosome on either side of the HC-contig/core centromeric DNA element to produce *in vivo* a stable HAC of minimal size.

Figure 15A shows an example of TACT-construct used in our study. Key features of this construct are: (a) Cloning of the pericentric human genomic DNA in both orientations (+/-). This is necessary since we do not know the chromosomal orientation of this DNA. This DNA is used to target the human telomeric sequences to locations on either side of the HC-contig region on mardel 10. Genomic DNA is derived from several different sources including Y2C24, Y3C64, Y3C109, Y3C94, Y13C12, Y13C15, Y17C6, Y17C8. The resulting truncation derivatives produced using these genomic DNAs will vary in size accordingly. (b) The termini contain 2.4 kilobases of tandem repeat human telomeric DNA (htel). This DNA has been shown previously to act as a substrate for mammalian telomerase to allow seeding of a complete telomere tens of kilobases in length. (c) The hygromycin (Hyg) resistance gene allows for positive selection of mammalian cell lines containing construct sequences integrated into the genome. This is the initial screening procedure. In addition, some constructs contain the neomycin phosophotransferase gene (Neo) rather than Hyg. (c) The Herpes simplex thymidine kinase (TK)

gene is used for negative selection against non homologous integration events into the genome. Those cell lines containing the TK gene can be selected against by adding the nucleoside analogue gancyclovir.

Figure 15B shows another example of TACT-construct used in our study. In addition to the features of the linearised construct shown in Fig. 15A, specific additional features are: (a) The incorporation of tandem telomeric blocks (htel.htel) since others have shown these to have the highest seeding efficiency of new telomeres in mammalian cells. (b) The incorporation of yeast selectable marker (eg. URA3), DNA origin of replication (eg. ARS), and centromere (eg. CEN6), to allow transfer and maintenance of the resulting truncation derivatives into yeast. This should facilitate further characterisation and manipulation, such as the introduction of therapeutic genes for gene therapy purposes. (c) The relocation of the TK gene adjacent to the genomic DNA to increase the effectiveness of the negative selection system. (d) The human growth hormone (GH) gene has been included to allow proof of principle that human genes can be introduced into a HAC and expressed under the control of endogenous regulatory elements. This is essential for gene therapy applications of the resulting HAC. (e) A CMV promoter upstream of a P1 phage loxP site (CMV/loxP) has been included to allow introduction of large human genes into a HAC *in vivo*. A plasmid containing a gene of interest, a second loxP site and a promoterless selectable marker gene is introduced into a mammalian cell line containing the HAC. Transient expression of CRE recombinase results in recombination between the two loxP sites within the cell, thereby integrating the introduced plasmid into the HAC and placing the selectable marker gene next to the CMV promoter to allow for marker selection.

For chromosomal truncation, the above TACT-constructs are transfected into a somatic cell hybrid line BE2CI-18-5f containing the mardel (10) chromosome. Positive selection is applied for Hygromycin or Geneticin resistance whereas negative selection is applied against the Thymidine Kinase Gene. Resulting colonies are further screened with distal p' and q' DNA fragments to ascertain the presence or absence of the two mardel 10 chromosome arms. In addition to the BE2CI-18-5f cell line, a human/chicken somatic cell hybrid line (derived from the recombination-proficient DT40 chicken cell line; Dieken *et al.*, 1996) containing the mardel (10) chromosome will also be generated and used.

EXAMPLE 16

Analysis of HAC

Irrespective of which of the approaches described above is used, the presence of a new product in a mammalian cell line as an extrachromosomal, artificial chromosome, will be assessed by fluorescence in situ hybridisation (FISH) analysis, as well as tested by extracting high molecular weight DNA to determine independently existing chromosomal entity on pulsed field gel. The stability of the construct through successive cell division, both in the presence and absence of drug-resistance selection, will be determined. The presence of the construct, in all or a high percentage of the original transfected cells indicates stability. Demonstration of this stability indicates the successful creation of a HAC.

EXAMPLE 17

Production of HAC

This example describes the use of the neocentromere as a source of centromeric DNA in the "bottom-up" approach to produce HACs in human cell culture. Bacterial artificial chromosomes (BACs) containing cloned neocentromeric DNA and a selectable marker were co-transfected with human telomeric DNA into human HT1080 cells to yield independent HACs that were single-copy and stable in the absence of selection. The properties of these HACs, and their potential utility as a new, improved vector system for gene therapy are described.

EXPERIMENTAL PROTOCOL

Preparation of DNA. Highly-purified BAC DNA was prepared using Qiagen columns according to the manufacturer's instructions. Prior to transfection, BACs were linearized with *Sgr*AI in the presence of 2.5 mM spermidine and examined by pulsed-field gel electrophoresis. Human telomeric DNA was gel-purified as a 1.6-kb *Bam*HI/*Bgl*II fragment from pSXneo270T2AG3 (Bianchi *et al.*, 1997). High-molecular-weight genomic DNA was prepared from cultured cell lines using standard methods (du Sart *et al.*, 1997).

Transfection of HT1080 cells. Transfection of human fibrosarcoma cell line HT1080 (Rasheed *et al.*, 1974) was performed using the DOPSER liposomal transfection reagent (Boehringer-Mannheim). The day before transfection, 6-well trays (each well is 962 mm^2) were seeded with 3×10^5 HT1080 cells per well and grown at 37°C, 5% CO_2. Different combinations containing 1-2 µg of each BAC, 50 ng of telomeric DNA, 100 ng of each PAC-1, 4 and 5 (du Sart *et al.*, 1997) and 50 ng of human genomic DNA were prepared in 50 µl of HBS (20 mM HEPES, 150 mM NaCl) supplemented with 0.075 mM spermidine and 0.030 mM spermine. These DNA cocktails were mixed with 50 µl of 0.4 µg/µl DOPSER (diluted in HBS) and left at room temperature for 15 to 20 min. The HT1080 cells were washed with PBS (phosphate buffered saline) and 1 ml of serum-free DMEM (Dulbecco's modified Eagles medium) was placed in each well. The DNA-DOPSER mixture was then added dropwise with swirling and the cells were incubated for 6 h. 1 ml of DMEM and 20% v/v fetal calf serum (FCS) was then added and the cells left for 24 h at 37°C, 5% v/v CO_2. The cells were harvested and seeded into 48-well cluster trays (each well is 100 mm^2) containing DMEM-10% v/v FCS supplemented with Geneticin (G418, Gibco-BRL) at 250 µg/ml. The media was changed every 3 to 4 days. G418-resistant colonies normally appeared 10 to 14 days after transfection. These colonies were expanded into duplicate 6-well trays, where the cells of one tray were stored frozen in liquid N_2, and the remaining cells were analysed by fluorescence *in situ* hybridization (FISH).

Cell culture and mitotic stability. HT1080 cells were grown in DMEM supplemented with 10% v/v FCS, penicillin/streptomycin, and glutamine. The mitotic stability of HAC containing clones was determined by growth in 25 cm^2 flasks in the presence (200-250 µg/ml) or absence of G418 selection, and grown to confluency (3-4 days) and split 1/5 and 1/10, respectively. Aliquots of each culture were harvested fortnightly and analysed by FISH (20-50 metaphases) with BAC/E8 and/or BAC/F2 probes.

FISH, ANTI-CEN/FISH and PRINS/FISH. Fluorescence *in situ* hybridization (FISH) analysis of HT1080 clones was performed with BAC/E8, BAC/F2, and/or α-satellite DNA probes. Hybridization using the BAC probes were performed under high stringency whereas the α-satellite DNA probes were used in low stringency conditions (du Sart *et al.*, 1997). ANTI-

CEN/FISH analyses involved an initial immunofluorescence staining step using a CREST antibody or specific antibodies against CENP-B, CENP-C, or CENP-E, followed by FISH using the probes described above, essentially as previously described (du Sart *et al.*, 1997).

Results

HAC construction strategy. The basic strategy involved the co-transfection of the 10q25.2 neocentromere DNA with human telomeric DNA into human cells. The neocentromere region is cloned as two, circular YACs in *Saccharomyces cerevisiae*. To facilitate handling and purification of the cloned DNA in large quantities, these YACs are retrofitted into BACs and maintained episomally in *E. coli* as circular molecules. One of the BAC clones, BAC/E8, is 120 kb in size and has an insert of 105 kb that encompassed 70 kb of the 80-kb core NC-DNA region (Fig. 16). The second BAC clone, BAC/F2, has an insert size of 75 kb that overlapped BAC/E8 by 1.4 kb, and contains ~10 kb of the core NC-DNA while extending ~65 kb into the p'-side of the mardel (10) chromosome (Fig. 16). The BAC vector backbone further contains the neomycin-resistance (Neo^R) gene to allow selection in mammalian cells. BAC/E8 and BAC/F2, used either on their own, in combination with each other or with additional DNA are used in the following transfection experiments.

Transfection of HT1080 cells. The human cell line HT1080 (Rasheed *et al.*, 1974) is chosen for the transfection experiments because of its near-diploid karyotype, its high level of telomerase activity (Holt *et al.*, 1997), and its demonstrated ability to form microchromosomes containing *de novo* centromeres from transfected arrays of α-satellite DNA and human telomeric DNA (Harrington *et al.*, 1997; Ikeno *et al.*, 1998). The resulting G418-resistant clones are analyzed by FISH and classified into different categories of events.

Transfected cell lines are designated HT-38, HT-47, HT-54, HT-190, and HT-191.

Those skilled in the art will appreciate that the invention described herein is susceptible to variation and modifications other than those specifically described. It is to be understood that the invention includes all such variations and modifications. The invention also includes all of the

steps, features, compositions and compounds referred to or indicated in this specification, individually or collectively, and any and all combinations of any two or more said steps or features.

TABLE 1

Restriction analysis of the genomic DNA of patient BE and those of his parents using three overlapping cosmids that span the marker centromere.

	Y7C14	Y6C10	Y4C7
NotI	n.a.	910	910
BssHII	n.a.	815, 340	n.a.
BsiWI	n.a.	740	740
SalI	410	410	410, 540
ClaI	315, 145, 110, 80	315, 145, 110, 80	315, 145, 110, 80
SnaBI	n.a.	250, 148	n.a.
NaeI	240, 210, 155, 120	240, 210, 155, 120	240, 210, 155, 120
NarI	222, 108, 70	222, 108	222, 200, 108, 70
EclXI	180	180	180
SfiI	170	170	170
KspI	168	168	168
AatII	165, 146	165, 146	165, 146
NheI	38	38	38
BstBI	n.a.	35	35
SmaI	n.a.	90, 40, 22	90, 40, 22
BglI	25	25, 7.2, 6.2	25
PacI	n.a.	25	n.a.
BamHI	24, 19, 15	24, 22*	24, 22*
KpnI	23	23	23, 19
BclI	21	21	21
PstI	9.4, 5.9, 5.1, 4.2, 3.8, 3.3, 2.9, 2.4	9.4, 3.8, 2.9, 2.7, 2.4, 1.5, 1.1	9.4, 7.1, 4.2, 3.3, 2.9, 2.7, 1.9, 1.5, 1.1
XbaI	14	14, 10	10
EaeI	n.a.	15, 12, 8, 6	n.a.
SphI	16, 7.5	16	16, 9
PvuII	14, 7.5	7.5, 6	7.5, 6
HindII	8.6, 6.9, 6.2, 2.7, 1.8, 1.2	6.9, 6.2, 5.6, 5.2, 5, 2.7, 1.9, 1.8, 1.7, 1.2, 0.6	6.2, 5.6, 5.2, 4.3, 2.9, 1.7, 1.2
ApaI	15, 8.5	15	15
EcoRI	11, 4.3, 3.9, 1.9, 1.5	11, 4, 3, 2, 1.9, 1.7, 1.5	10.2, 7.6, 3, 2, 1.9, 1.7, 1.5
HpaII	5.5, 4.3, 3.6, 1.6	6.9, 3.6, 2.8, 1.6, 1.2	3.6, 2.8, 2.5, 1.6, 1.2

MspI	3.9, 3.0, 2.8, 2.5, 2, 1.6, 1.2	3.9, 3.6, 2.8, 2.5, 2.2, 1.6, 1.5, 1.3, 1.2, 0.9	3.6, 3.2, 2.8, 2.5, 2.2, 1.6, 1.5, 1.2, 1
SspI	n.a.	10	n.a.
XhoII	7.5	n.a.	n.a.
DraI	7.5	7.5	7.5
BglII	8.5, 6, 5, 4.7, 3.5, 2.5	6, 5, 4.7, 2.5, 1.6, 1.5, 1	7, 6, 5, 4.7, 2.5, 1.6, 1.5, 1.1, 1
AvaII	7.4, 3.7, 3.4, 2.8, 2.6, 1.8, 1.7, 1.4, 1.2, 1.1	3.7, 2.8, 2.6, 1.8, 1.7, 1.4, 1.2, 1.1, 0.9, 0.8, 0.5	4.3, 3.7, 2.8, 2.6, 1.8, 1.7, 1.4, 1.2
StuI	12.5, 8, 7.5	12.5, 9, 8.5	9, 8.5
HindIII	6.6, 5.4, 4.7, 4.4, 2.9, 2.5	5, 4.7, 4.4, 4.1, 2.9, 2.5, 0.7	5, 4.7, 4.1, 3.1, 2.5, 2.3, 1.9

n.a.= data not available.The values represent restriction fragment lengths in kilobases. Multiple values for an enzyme denote different bands detected by a cosmid probe on a gel lane. Since there were no detectable differences between the DNA of patient BE and those of his parents in any of the fragments (except for a BamHI polymorphic band found in one of the parents, indicated by an asterisk), only one set of values is shown for all three genomic DNA.

TABLE 2

Table 2. Vectors for cloning centromeric regions from normal chromosome 10 or mardel (10) DNA into yeast artificial chromosomes (YACs). These YACs can be shuttled into mammalian cells to test for function as HACs.

Vector:	Key Feature(s)	Map
pJS97ARTi	hTEL/I-SceI/yTEL, DHFR	Fig.8A
pJS98ANTi	hTEL/I-SceI/yTEL, neo	Fig.8B
Fragmentation 1	hTEL/I-SceI/yTEL, hyg	Fig.8C
Fragmentation 2 (-/+ hGH)	hTEL/I-SceI/yTEL, neo, hGH	Fig.8D
pVC39-AAH2	ALU-ALU TAR vector	Fig.8E
pTEL/CAT/TEL	hTEL/I-SceI/hTEL/neo	Fig.8F
pAAH/TCNa	TAR vector with hTEL/I-SceI/hTEL/neo	Fig.8G
pVC39-ALU/C3-F2(+/-)	ALU-specifc TAR vectors	Fig.8H
pTCS	ends of PAC4 in pBS	Fig.8I
pTCN-TCS	specific TAR vector hTEL/I-SceI/hTEL/neo	Fig.8J

TABLE 3

PCR analysis of YAC 5f-52-E8 clone and comparison with the HC-contig/neo-centromere region from normal chromosome 10 and mar del (10)

Primer-Pairs [a]	Genomic DNA used in PCR (product size in kb)		
	BE2C1-18-1f[b]	BE2C1-18-5f[b]	YAC 5f-52-E8
norm: 141 + 55	1.80	1.80	not present
norm: 32 + 30	0.90	0.90	0.90
norm: 28 + 29	1.00	1.00	1.00
norm: 1 + 3	2.90	2.90	2.90
norm: 39 + 52	1.20	1.20	1.20
norm: 5 + 7	0.23	0.23	0.23
norm: 16 + 5	3.50	3.50	3.50
norm: 9+ 14	0.90	0.90	0.90
norm: 36 + 37	2.00	2 00	2 00
norm: 168 + 71	4.00	4.00	4.00
norm: 27 + 10	15.90	15.90	15.90
norm: 18 + 17 (VNTR)[c]	1.20	1.40	1.40
norm: 68 + 17	8.00	8.00	8.00
norm: 34 + 47	3.00	3.00	3.00
PAC4t7: a + b	0 30	0.30	not present
AFM259xg5: ca + gt [c]	0.21	0.19	not present

[a] Refer to Fig. 1a for the relative positions of each primer-pair.

[b] BE2C1-18-1f and BE2C1-18-5f are somatic hybrid cell lines containing the normal human chromosome 10 and mar del (10), respectively (2).

[c] The 'norm: 18 + 17' and 'AFM259xg5: ca and gt' primer sets allow distinction between the normal human chromosome 10 and mar del (10) by detecting a VNTR and a microsatellite, respectively.

BIBLIOGRAPHY:

1. Albertsen, H., Abderrahim, H., Cann, H., J, D., Paslier, D. L., and Cohen, D. (1990). Construction and characterization of a yeast artificial chromosome library containing seven haploid human genome equivalents. *Proc. Natl. Acad. Sci. USA. 87*, 4256-4260.
2. Archidiacono, N., Antonacci, R., Forabosco, A., and Rocchi, M. (1994). Preparation of human chromosomal painting probes from somatic cell hybrids. In *In Situ* Hybridization Protocols. K. H. A. Choo, ed. (Totowa, New Jersey: Humana Press), pp. 1-14.
3. Bernat, R. L., Borisy, G. G., Rothfield, N. F., and Earnshaw, W. C. (1990). Injection of anticentromere antibodies in interphase disrupts events required for chromosome movement in mitosis. *J. Cell. Biol. 111*, 1519-1533.
4. Bischoff, F., Maier, G., Tilz, G., and Ponstingl, H. (1990). A 47-kDa human nuclear protein recognized by antikinetochore autoimmune sera is homologous with the protein encoded by RCC1, a gene implicated in onset of chromosome condensation. *Proc. Natl. Acad. Sci. 87*, 8617-8621.
5. Brenner, S., Pepper, D., Berns, M. W., Tan, E., and Brinkley, B. R. (1981). Kinetochore structure, duplication and distribution in mammalian cells: analysis by human autoantibodies from scleroderma patients. *J. Cell. Biol. 91*, 95-102.
6. Brown, K. E., Barnett, M. A., Burgtorf, C., Shaw, P., Buckle, V. J., and Brown, W. R. A. (1994). Dissecting the centromere of the human Y chromosome with cloned telomeric DNA. *Hum. Mol. Genet. 3*, 1227-1237.
7. Brownstein, B., Silverman, G., Little, R., Burke, D., Korsmeyer, S., Schlessinger, D., and Olson, M. (1989). Isolation of single-copy human genes from a library of yeast artificial chromosome clones. *Science 244*, 1348-1351.
8. Clarke, L., and Carbon, J. (1985). The structure and function of yeast centromeres. *Annu. Rev. Genet. 19*, 29-56.
9. Dasso, M. (1993). RCC1 in the cell cycle: the regulator of chromosome condensation takes on new roles. *Trends Biochem. Sci. 18*, 96-101.
10. Dieken *et al.* (1996) *Nature Genetics 12:* 174-182.
11. du Sart, D., Cancilla, M.R., Earle, E., Mao, J., Saffery, R., Tainton, K. M., Kalitsis, P., Martyn, J., Barry, A.E., and Choo, K.H.A. (1997). A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellit DNA. *Nature Genet. 16*, 144-153.

12. du Sart, D., Cancilla, M.R., Earle, E., Mao, J., Saffery, R., Tainton, K.M., Kalitsis, P., Martyn, J., Barry, A.E., and Choo, K.H.A. 1997. A functional neo-centromere formed through activation of a latent human centromere and consisting of non-alpha-satellite DNA. *Nature Genetics* **16:**144-153.

13. Harrington, J.J., Van Bokkelen, G., Mays, R.W., Gustashaw, K., and Willard, H.F. 1997. Formation of *de novo* centromeres and construction of first-generation human artificial microchromosomes. *Nature Genetics* **15:**345-355.

14. Holt, S.E., Aisner, D.L., Shay, J.W., and Wright, W.E. 1997. Lack of cell cycle regulation of telomerase activity in human cells. *Proc. Natl. Acad. Sci. USA* **94:**10687-10692.

15. Ikeno, M., Grimes, B., Okazaki, T., Nakano, M., Saitoh, K., Hoshino, H., McGill, N.I., Cooke, H., and Masumoto, H. 1998. Construction of YAC-based mammalian artificial chromosomes. *Nature Biotechnology* **16:**(in press).

16. Earnshaw, W., and MacKay, A. (1994). Role of nonhistone proteins in the chromosomal events of mitosis. *FASEB J. 8*, 947-956.

17. Earnshaw, W. C., and Migeon, B. R. (1985). Three related centromere proteins are absent from the inactive centromere of a stable isodicentric chromosome. *Chromosoma 92*, 290-296.

18. Earnshaw, W. C., Ratrie, H., and Stetten, G. (1989). Visualization of centromere proteins CENP-B and CENP-C on a stable dicentric chromosome in cytological spreads. *Chromosoma 98*, 1-12.

19. Farr, C., Bayne, R., Kipling, D., Mills, W., Critcher, R., and Cooke, H. (1995). Generation of a human X-derived minichromosome using telomere-associated chromosome fragmentation. *EMBO Journal 14*, 5444-5454.

20. Fritzler, M. J., and Kinsella, T. D. (1980). The CREST syndrome: a distinct serologic entity with anticentromere antibodies. *Am. J. Med. 69*, 520-526.

21. Grady, D., Ratliff, R., Robinson, D., McCanlies, E., Meyne, J., and Moyzis, R. (1992). Highly conserved repetitive DNA sequences are present at human centromeres. *Proc. Natl. Acad. Sci. USA 89*, 1695-9.

22. Haaf, T., and Ward, D. C. (1994). Structural analysis of α-satellite DNA and centromere proteins using extended chromatin and chromosomes. *Hum. Mol. Genet. 3*, 697-709.

23. Haaf, T., Warburton, P. E., and Willard, H. F. (1992). Integration of human α-satellite DNA into simian chromosomes: centromere protein binding and disruption of normal

chromosome segregation. *Cell 70*, 681-696.

24. Jeppensen, P., Mitchell, A., Turner, B., and Perry, P. (1992). Antibodies to defined histone epitopes reveal variations in chromatin conformation and underacetylation of centric heterochromatin in human metaphase chromosomes. *Chromosoma 101*, 322-332.

25. Jeppensen, P., and Turner, B. M. (1993). The inactive X chromosome in female mammals is dinstinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. *Cell 74*, 281-289.

26. Kingwell, B., and Rattner, J. (1987). Mammalian kinetochore/centromere composition: A 50 kDa antigen is present in the mammalian kinetochore/centromere. *Chromosoma 95*, 403-407.

27. Larin, Z., Fricker, M. D., and Tyler-Smith, C. (1994). De novo formation of several features of a centromere following introduction of a Y alphoid YAC into mammalian cells. *Hum. Mol. Genet. 3*, 689-695.

28. Larionov, V. *et al.* (1997) *Proc. Natl. Acad. Sci. USA 94:* 7384-7387.

29. Larionov, V., Kouprina, N., Graves, J., Chen, X. N., Korenberg, J. R., and Resnick, M. A. (1996a). Specific cloning of human DNA as yeast artificial chromosomes by transfromation-associated recombination. *Proc. Nat. Acad. Sci. USA* **93**, 491-496.

30. Larionov, V., Kouprina, N., Graves, J., and Resnick, M. A. (1996b). Highly selective isolation of human DNAs from rodent-human hybrid cells as circular yeast artificial chromosomes by transformation-associated recombination cloning. *Proc. Nat. Acad. Sci. USA* 93, 13925-13930.

31. Moir, D. T., Dorman, T. E., Day, J. C., Ma, N. S., Wang, M., and Mao, J. (1994). Toward a physical map of human chromosome 10: isolation of 183 YACs representing 80 loci and regional assignment of 94 YACs by fluorescence in situ hybridization. *Genomics 22*, 1-12.

32. Moroi, Y., Hartman, A. L., Nakane, P. K., and Tan, E. M. (1981). Distribution of kinetochore (centromere) antigen in mammalian cell nuclei. *J. Cell Biol. 90*, 254-259.

33. Moschonas, N. K., Spurr, N. K., and Mao, J. (1996). Report of the first international workshop on human chromosome 10 mapping 1995. *Cytogenet. Cell Genet. 72*: 99-112.

34. Murphy, T. D., and Karpen, G. H. (1995). Localization of centromere function in a Drosophila minichromosome. *Cell 82*, 599-609.

35. Nelson, M., and McClelland, M. (1991). Site-specific methylation: effect on DNA modification methyltransferases and restriction endonucleases. Nucl. Acids Res. *19*:

2045-2071.

36. Page, S. L., Earnshaw, W. C., Choo, K. H. A., and Shaffer, L. G. (1995). Further evidence that CENP-C is a necessary component of active centromeres: studies of a dic(X;15) with simultaneous immunofluorescence and FISH. *Hum. Mol. Genet. 4*, 289-294.

37. Pluta, A. F., Cooke, C. A., and Earnshaw, W. C. (1990). Structure of the human centromere at metaphase. *Trends Biochem. 15*, 181-185.

38. Pluta, A. F., Mackay, A. M., Ainsztein, A. M., Goldberg, I. G., and Earnshaw, W. C. (1995). The centromere: hub of chromosomal activities. *Science 270*, 1591-1594.

39. Rasheed, S., Nelson-Rees, W.A., Toth, E.M., Arnstein, P., and Gardner, M.B. (1974) Characterisation of a newly derived human sarcoma line (HT1080). *Cancer* **33**, 1027-1033.

40. Sikorski, R.S. and Hieter, P. (1989). A system of shuttle vectors and yeast host strains designed for efficient manipulation of DNA in *Saccharomyces cerevisiae. Genetics 122*, 19-27.

41. Steiner, N., Hahnenberger, K., and Clarke, L. (1993). Centromeres of the fission yeast Schizosaccharomyces pombe are highly variable genetic loci. *Mol. Cell. Biol. 13*, 4578-4587.

42. Sullivan, B.A. and Schwartz, S. (1995). Identification of centromeric antigens in dicentric Robertsonian translocations: CENP-C and CENP-E are necessary components of functional centromeres. Hum. Mol. Genet. *4*, 2189-2197.

43. Sullivan, K. F., Hechenberger, M., and Masri, K. (1994). Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere. *J. Cell Biol. 127*, 581-592.

44. Taylor, S.S., Larin, Z., and Tyler-Smith, C. (1994) Addition of functional human telomeres to YACs. *Human Mol Genet* **3**, 1383-1386.

45. Taylor, S.S., Larin, Z., and Tyler-Smith, C. (1996) Analysis of extrachromosomal structures containing human centromeric alphoid satellite DNA sequences in mouse cells. *Chromosoma* **105**, 70-81.

46. Tomkiel, J., Cooke, C. A., Saitoh, H., Bernat, R. L., and Earnshaw, W. C. (1994). CENP-C is required for maintaining proper kinetochore size and for a timely transition to anaphase. *J. Cell. Biol. 125*, 531-545.

47. Trowell, H. E., Nagy, A., Vissel, B., and Choo, K. H. A. (1993). Long-range analyses

of the centromeric regions of human chromosomes 13, 14 and 21: identification of a narrow domain containing two key centromeric DNA elements. *Hum. Mol. Genet. 2*, 1639-1649.

48. Tyler-Smith, C., Oakey, R. J., Larin, Z., Fisher, R. B., Crocker, M., Affara, N. A., Ferguson-Smith, M.A., Muenke, M., Orsetta, Z., and Jobling, M. A. (1993). Localization of DNA sequences required for human centromere function through an analysis of rearranged Y chromosomes. *Nature Genet. 5*, 368-375.

49. Voullaire, L. E., Slater, H. R., Petrovic, V., and Choo, K. H. A. (1993). A functional marker centromere with no detectable alpha-satellite, satellite III, or CENP-B protein: activation of a latent centromere. *Am. J. Hum. Genet. 52*, 1153-1163.

50. Wevrick, R., and Willard, H. F. (1989). Long-range organization of tandem arrays of alpha-satellite DNA at the centromeres of human chromosomes: high-frequency array-length polymorphism and meiotic stability. *Proc. Natl. Acad. Sci. USA 86*, 9394-9398.

51. Wevrick, R., and Willard, H. F. (1991). Physical map of the centromeric region of human chromosome 7: relationship between two distinct alpha satellite arrays. *Nucl. Acids Res. 19*, 2295-2301.

52. Zheng, C., Ma, N. S., Dorman, T. E., Wang, M., Braunschweiger, K., Soares, L., Schuster, M. K., Rothschild, C. B., Bowden, D. W., Torrey, D., Keith, T. P., Moir, D. T., and Mao, J. (1994). Development of 124 sequence-tagged sites and cytogenetic localization of 217 cosmids for human chromosome 10. *Genomics 22*, 55-67.